ABSTRACT

Title of dissertation: ESTIMATION AND ADAPTIVE ONLINE CORRECTION OF SYSTEMATIC ERRORS IN THE GLOBAL FORECAST SYSTEM (GFS) USING ANALYSIS INCREMENTS

Kriti Bhargava
Doctor of Philosophy, 2019

Dissertation directed by: Professor Eugenia Kalnay and
Professor James A. Carton
Department of Atmospheric and Oceanic Science

Numerical Weather prediction models have improved drastically in the last few decades with advances in data assimilation, improved parameterizations, and ensemble forecasting. Despite these developments, the performance of numerical weather prediction models like the Global Forecast System (GFS) are still limited by errors in the model forecasts. These errors arise from inaccuracies in the initial condition and models inability to accurately represent physics, dynamics and chemical processes. Operation centers generally use an offline correction schemes that corrects the forecast error after the forecast is generated. Past research has shown that another class of correction schemes, the online correction schemes that correct for the forecast errors during the model integration have certain advantages over offline schemes. However, the online schemes tested so far are prohibitive for operation use. The goal of this work is to introduce and test an "adaptive online correction scheme based on the methodology developed by Danforth et al. (2007)

that is suitable for operational use is introduced and implemented.

As a first step towards correcting tendency equation, the model errors are estimated using the 6-hr Analysis Increments (AIs). Assuming initial linear error growth and absence of any residual bias in the analysis, 6-hr AIs provide a measure of model errors that can later be used to estimate model tendency errors. Seasonal means of 6-hr AIs during the period from 2012-2016 indicate robust model biases despite the changes in the model and data assimilation during that period. Apart from the season mean error, GFS also has significant periodic errors that are dominated by errors in the diurnal and semi-diurnal cycle.

An adaptive online correcting scheme that uses 6-hr AIs, averaged over a moving training period to compute the bias correction term to be added in model integration equation is then implemented with GFS. The scheme is tested using training periods of different lengths ranging from past 7 to 28 days. This scheme is remarkably stable and reduces the forecasts errors significantly in forecasts all over the globe at lead times of 1 day and shorter and over the tropics at longer lead times. An offline correction scheme was also tested but found to be less effective than the online correction scheme especially at lead times longer than 1-day.

ESTIMATION AND ADAPTIVE ONLINE CORRECTION OF SYSTEMATIC ERRORS IN THE GLOBAL FORECAST SYSTEM (GFS) USING ANALYSIS INCREMENTS

by

Kriti Bhargava

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Distinguished University Professor Eugenia Kalnay, Chair/Advisor
Professor James A. Carton, Co-Advisor
Adjunct Assistant Professor Daryl Kleist
Assistant Research Professor Stephen G. Penny
Professor and Dean's Representative Brian R. Hunt

Dedication

This dissertation is dedicated to my very hardworking and beloved grandparents, Bijendra Nath and Shanti Bhargava; and T.P. and Nirmala Bhargava, for their unconditional love and strong values.

My dedication also goes to many great teachers and advisors who have guided me throughout my journey from learning "abc" to earning a "Dr.' before my name.

Acknowledgments

I owe tremendous gratitude to all the people who have made this thesis possible and because of whom my graduate experience has been one that I will cherish forever. Firstly, I wish to express my deepest gratitude to my advisors, Dr. Eugenia Kalnay and Dr. Carton for providing me the opportunity to work in the field of Data assimilation and Numerical Weather Prediction despite my lack of prior background. I consider it a privilege to have them as my advisors and thank them immensely for guiding and supporting me though the entire period of my graduate study. I thank Dr. Kalnay for her brilliant ideas, patience during the spells of slow progress, consistent encouragement and praises that kept me motivated and boosted my confidence and devoting her valuable time reviewing and helping me with my writing. I also thank her for introducing me to several scientists who were crucial in running the GFS model. I thank Dr. Carton for being understanding and supportive both mentally and academically, providing hand on session on concepts I struggled with, encouraging me to speak up and largely for his valuable advice, time and efforts helping me improve my writing skills.

Next I would like to thank my parents specially my mother, Charu who was my first teacher and instilled in me the love of science and mathematics and has always been my strong advocate. I am also very grateful to my brother, Apoorv who very diligently proofread all my documents, kept me motivated. I am tremendously grateful to my husband, Harshit for his undying faith in me, support and encouragement. I thank my brother and husband for cooking and making sure I ate

properly while I was writing my thesis. I am also very grateful for my supporting and understanding in-laws.

I thank Dr. Fanglin yang for kindly providing the 2012-2016 GFS analysis and forecast datasets. I am very grateful to Dr. Jim Jung for his valuable time and willingness to help with running GFS on the S4 supercomputer. I am also grateful to Dr. Mark Iredell, Henry Juang and Srinivas Moorthi for helping decode the very involved GFS code.

I am also thankful to my committee members Drs. Brian Hunt, Daryl Kleist and Steve Penny for their guidance, discussions, feedbacks and insights that have added greatly added to my understanding and this dissertation.

I feel deep gratitude towards Dr. Sumant Nigam who acted as a mentor throughout my graduate study and was always available for advice and discussions. Transition into the academic environment of a new country and new discipline would have have been much harder without him.

I also thank Adria, Tse-Chun, Travis Sluka, Rebekkah, Erin Lynch, Cheng-Da and several fellow students for their insights and feedback.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

3DVAR	3-D Variational
6-hr	Six Hours
AI	Analysis Increment
AOC	Adaptive Online Correction
CART-ACE	Classification and Regression trees and the Alternative Conditional Expecta
CRTM	Community Radiative Transfer Model
DA	Data Assimilation
DJF	December January February
ECMWF	European Center for Medium-Range Forecast
ENKF	Ensemble Kalman Filter
EOF	Empirical Orthogonal Function
ERA5	ECMWF Reanalysis 5
GCM	Global Circulation Model
GDAS	Global Data Assimilation System
GFS	Global Forecast System
GRAPES	Global and Regional Assimilation and Prediction System
GSI	Gridpoint Statistical Interpolation
ITCZ	Inter Tropical Convergence Zone
JJA	June July August
LDW	Linearly Decaying Weight
MAM	March April May
MOS	Model Output Statistics
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
OI	Optimally Interpolated
OSSE	Observation System Simulation Experiments
NMC	National Meteorological Center
NNR	NCEP-NCAR Reanalysis
NSST	Near-Surface Sea Temperature
NWP	Numerical Weather Prediction
RHS	Right Hand Side
RMS	Root Mean Square
RMSE	Root Mean Square Error
RTG	Real Time Global
SFE	Systematic Forecast Error
SON	September October November
SPEEDY	Simplified Parameterizations primitivE Equation Dynamics
SST	Sea Surface Temperature
SVD	Singular Value Decomposition

Chapter 1: Introduction

1.1 Errors in Numerical Weather Forecast Model

Numerical weather prediction (NWP) has evolved significantly since the early 1900s, when it was proposed by Bjerknes (1904) as a scientific initial value problem (Kalnay, 2003). Bauer et al. (2015) have shown that forecast skill of short to medium range forecast has been improving by one day per decade since the past 40 years1.1. This improvement in NWP is accredited mainly to improvement in (1) physical process representation: better understanding of physics, improved physical parameterizations, more accurate and stable numerical discretization schemes, and increase in computing resources; (2) model initialization: more observations, advancement in data assimilation; and (3) ensemble forecasting (Houtekamer et al., 2016; Hunt et al., 2007; Ott et al., 2002).

Despite these advancements, performance of NWP models is still largely limited by errors in the model forecasts. WMO-exchanged scores from global forecast centers, including Japan Meteorological Agency, Canadian Meteorological Centre, UK Met Office, U.S. National Centers for Environmental Prediction, Mto France, Deutscher Wetterdienst, and European Center for Medium-Range Weather Forecasts, show that the models have improved in terms of predicting 500hPa height

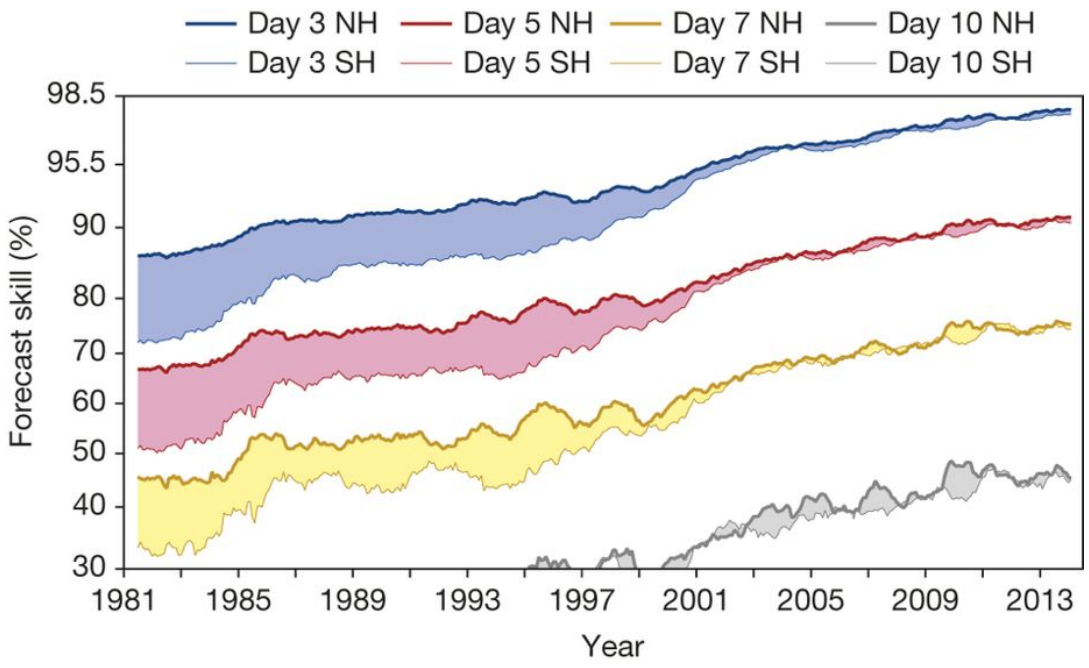

Figure 1.1: Improvement in the forecast skill since 1980s. Forecast skill is the correlation between the forecasts and the verifying analysis of the height of the 500-hPa level, expressed as the anomaly with respect to the climatological height. Values greater than 60% indicate useful forecasts, while those greater than 80% represent a high degree of accuracy. (Bauer et al., 2015)

(Fig. 1.2). However, significant systematic errors still exist that grow as the model is integrated further in time (Fig. 1.3).

The reasons for these errors are similar to the factors responsible for improvements in NWP. As NWP is an initial/boundary value problem, a major source of forecast error is inaccuracy in initial conditions. The other source of error is model's inability to accurately represent physics, dynamics and chemical processes. Errors in the initial conditions are amplified by atmospheric instabilities with time, also referred to as internal error or displacement error growth (Kalnay, 2003; Orrell et al., 2001). Model deficiencies arise due to inaccurate formulation and numerical discretization of the equations of motion, imperfect parameterizations that represent the effect of sub-grid processes and boundary conditions. These model deficiencies introduce errors, also referred to as model drift, and lead to "external" error growth as the model is integrated with time. In addition to worsening the forecast, these model errors can also lead to divergence of data assimilation schemes (Li et al., 2009). Most data assimilation schemes assume forecast and observations are unbiased and account only for the known variance. In the presence of forecast bias, the DA system fails to minimize error variance accurately and result in biased analysis fields.

Accurate and timely forecasts not only help save lives and help in preparedness in case of extreme weather but also impact several sectors of economy. Sectors of economy largely impacted by weather forecasts include energy, agriculture, insurance, telecommunication, tourism, transport (aviation, road, and railway) and shipping. Though it is hard to accurately quantify the socio-economic benefits of

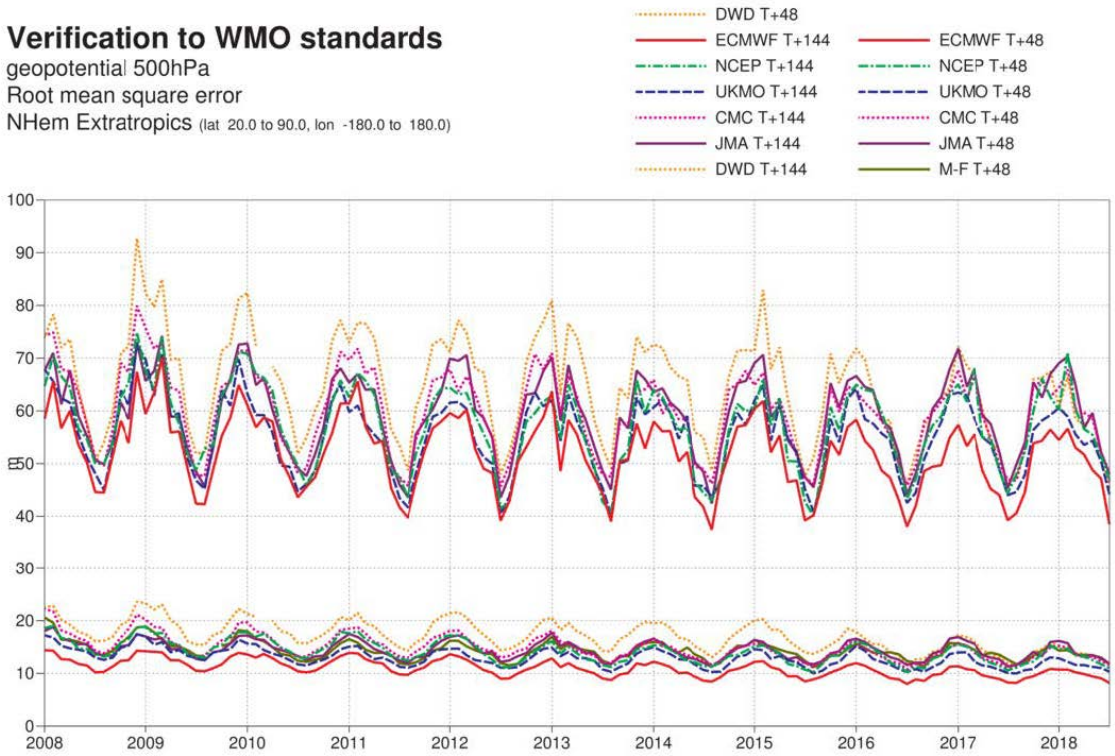

Figure 1.2: WMO-exchanged scores from global forecast centres. RMS error of 500 hPa geopotential height over northern extratropics. Upper curves show the six-day forecast error and the lower curves show the two-day forecast error of model runs initiated at 12 UTC. Each model is verified against its own analysis. JMA = Japan Meteorological Agency, CMC = Canadian Meteorological Centre, UKMO = the UK Met Office, NCEP = U.S. National Centers for Environmental Prediction, M-F = Mto France, DWD = Deutscher Wetterdienst. (Source:(Haiden, T, Janousek, M, Bidlot, J-R, Buizza, R, Ferranti, L, Prates, F, Vitart, 2018))

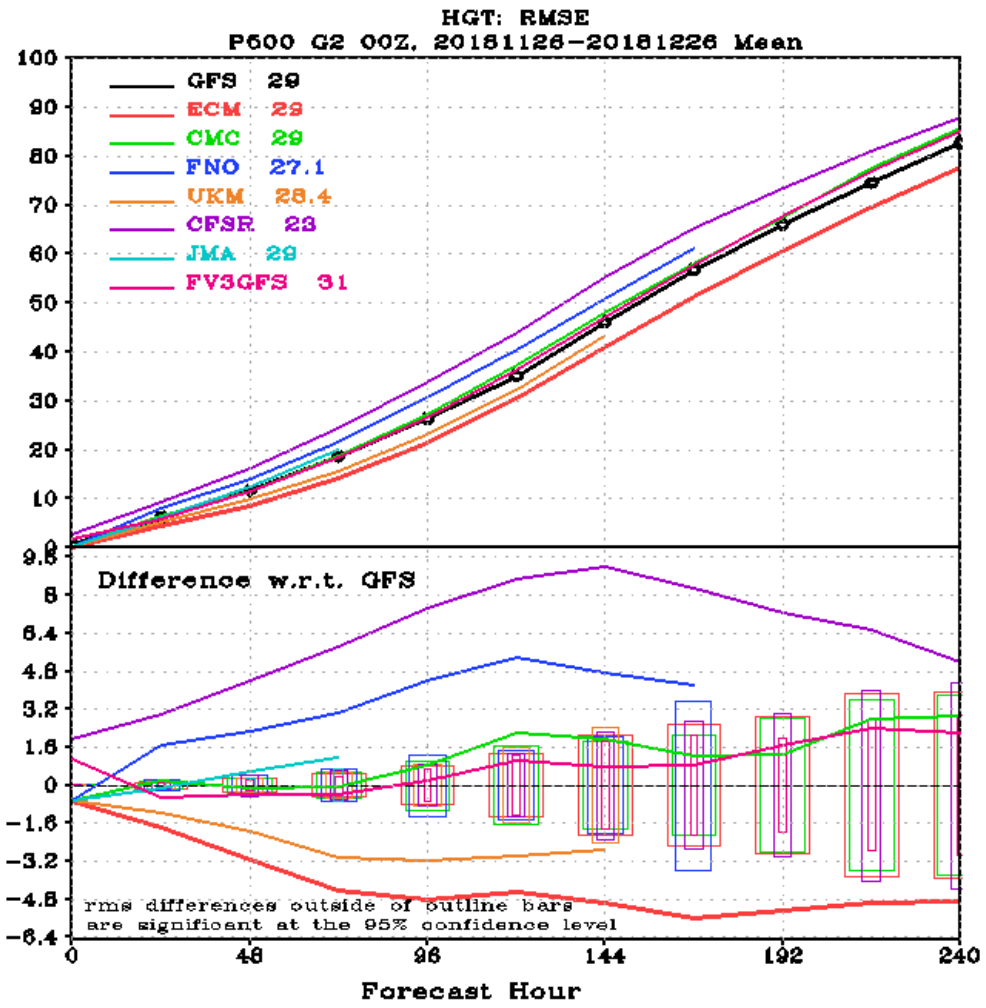

Figure 1.3: RMS error of 500hPA geopotential height over whole globe verified against GFS analysis. Model Acronyms: GFS: The U.S. NCEP Global Forecast Systems; ECM: European Center for Medium-Range Weather Forecasts; CMC: The Canadian Meteorological Center; FNO: The U.S. Navy Fleet Numerical Meteorology and Oceanograpy Center; UKM: The United Kingdom Met Office; JMA: Japan Meteorological Agency; CFSR: Legacy GFS used for Climate Forecast System Reanalysis (Source: https://www.emc.ncep.noaa.gov/gmb/STATS_vsdb/)

Country	economic sector	cost-benefit ratio	Reference
Australia	general public (willingness to pay analysis) in Sydney	1 : 4	Anamann and Lellyett (1996)
Croatia	overview of all sectors	1 : 3 (at least)	Leviäkangas et al. (2008)
Denmark	overview of all sectors, 3 case studies	Many detailed examples with good net benefits; aggregate picture lacking	The Ministry of Transport and Energy (2006)
Finland	overview of all sectors	1 : 5	Leviäkangas and Hautala (2009)
	transport	1 : 10 and higher	Nurmi et al. (2012) internal FMI study (on-going for other sectors)
Nepal	mainly agriculture; transport and hydropower also considered	around 1 : 10	Perrels (2011)
Russia	overview of all sectors	1 : 3–1 : 4	Bedritsky and Khandozko (2001)
Switzerland	overview, all sectors	1 : 5	Frei (2010)
	transport sector	1 : 10	Frei et al. (2012)
United Kingdom	general public (willingness to pay analysis)	1 : 7	Met Office (2007)
	meteorological infrastructure – satellite	1 : 5–1 : 20	Joo et al. (2011)
USA	transport sector (winter road maintenance)	1 : 2–1 : 3	Ye et al. (2009)
	general public (willingness to pay analysis)	1 : 6	Lazo et al. (2009)

Figure 1.4: Overview of published cost-benefit analysis within different weather services, based on avoided costs unless otherwise stated. (Source: Perrels et al. (2013))

NWP, several studies have attempted estimating cost to benefit ratio for improving the NWP models (Frei, 2009; Perrels et al., 2013). All these studies indicate very favorable cost to benefit ratio. These estimates are based on several assumptions, different methodologies and differ for different countries and different sectors. Perrels et al. (2013), summarized the results of these studies as given in fig 1.4

In order to improve the forecasts, it is important to continue improving the initial conditions by assimilating observations and ensemble forecasts along with

reducing model deficiencies by improving the physical parameterizations, increasing resolution to resolve smaller-scale processes and generating more accurate forcings. However, even with the most accurate and highest resolution grids, models would still contain errors due to unresolved phenomenon and smaller state space compared to true nature. Additionally, despite knowing possible areas that need improvements, it remains one of the major challenges in NWP to identify the physical sources of model error in complex NWP models. Hence, it is important to develop empirical schemes that aim at reducing forecast errors arising from model deficiencies.

This dissertation focuses on estimating and correcting the systematic component (mentioned in next section and discussed in further detail in Chapter 2) of model forecast error in the Global Forecast System (GFS). The GFS is a global numerical weather prediction model which provides 16-day forecasts produced by the National Centers for Environmental Prediction (NCEP). It couples an atmosphere model, and a land/soil model and uses near-surface sea temperature (NSST) to provide realistic ocean boundary conditions to produce the forecasts. It is initialized with the Global Data Assimilation System (GDAS). Details about GFS are discussed in Chapter 2 and Chapter 4.

1.2 Systematic Errors in GFS

One major problem that arises in the discussion about model errors is that it means different things to different people (Allen et al., 2006; Dalcher and Kalnay,

1987; Orrell et al., 2001). The most commonly used performance metric to evaluate model performance and quantify model forecast error has been the Root Mean Square (RMS) error. Yang (2015) showed that RMSE at times fails to provide an accurate measure of model performance. Dalcher and Kalnay (1987), taking the advantage of additive property of variance, separated the mean square error (eqn. 1.1) into the (1) square of differences of mean, referring to it as the systematic components and (2) and mean square difference of standard deviations, referred to as non-systematic components. In eqn. (1.1) x_t refers to the verifying truth, x_f refers to the forecast, overbar represents spatial/temporal mean and the primes represent the deviation from the respective mean state. The unbiased second component is the error due to pattern variation and apart from depending on standard deviations also depends on the anomalous pattern correlation (Yang, 2015).

$$\overline{(x_f - x_t)}^2 = (\overline{x_f} - \overline{x_t})^2 + \overline{(x_f' - x_t')}^2 \tag{1.1}$$

Systematic forecast errors (SFEs), as defined by eqn. 1.1, form a significant portion of the total forecast error in weather prediction models, such as the Global Forecast System (GFS). Fig. 1.5 shows that after two weeks, the range of GFS RMS temperature systematic errors reaches about one third of the total temperature forecast error range. Many studies attribute SFEs to specific deficiencies in numerical discretization of the equations of motion, parameterizations of subgrid scale processes, or boundary conditions (Allen et al., 2006; Jung, 2005; Jung and Tompkins, 2003b), which lead to model bias. These errors are initially small, but as

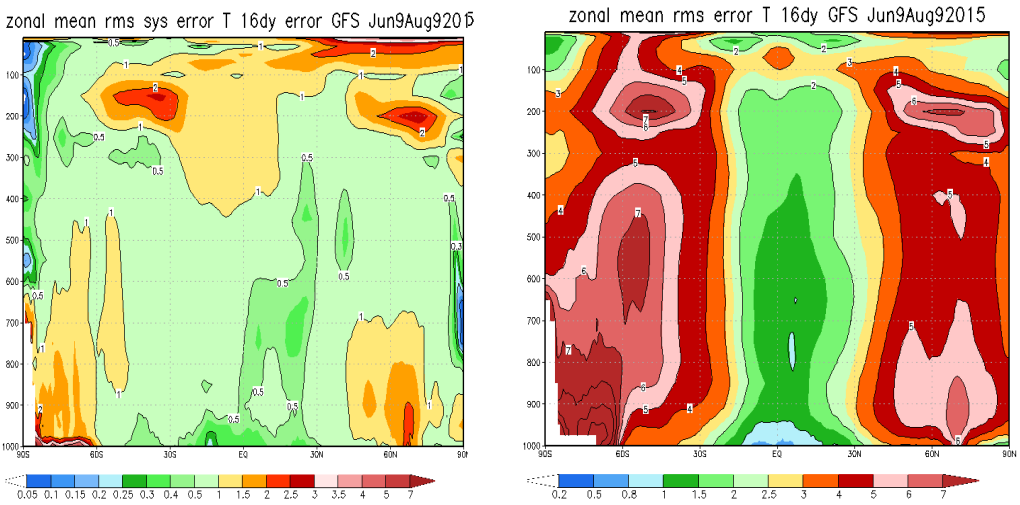

Figure 1.5: Zonal mean RMS (left) systematic error and (right) total error in temperature after 16 days. The range of temperature systematic errors is approximately one third of total temperature error range after 2 weeks (courtesy of Dr. Glenn White).

the model is integrated in time the errors grow and interact non-linearly with other systematic and random errors until the model loses all forecast skill.

SFEs also include periodic forecast errors, including those associated with the annual cycle and the diurnal cycle, and also state-dependent errors that are highly correlated with certain states of atmosphere. For example, systematic errors arise from the presence of weather highs and lows, or the phase of El-Niño (Danforth et al., 2007). According to the definition in equation 1.1, periodic and state-dependent errors would erroneously be considered as random. This dissertation addresses that problem in Chapter 2, where the systematic and random errors are redefined.

1.3 Empirical Correction Schemes

Several empirical schemes have been proposed and tested to correct systematic errors in NWP models. These schemes can be classified as (1) offline schemes, also known as post-processing techniques, where an empirical correction is applied after the forecast have been made, and, alternatively, (2) online schemes where the model tendency equations are adjusted to account for model systematic errors.

1.3.1 Offline schemes: Brief Review

Operational NWP models generally use offline approaches for the removal of systematic errors. A widely used operational approach is the Model Output Statistics (MOS) approach (Carter et al., 1989; Glahn and Lowry, 1972). MOS is based on multiple linear regression utilizing the correlation between the model forecast (pre-

dictor) and observed values (predictand). One major disadvantage of MOS is that it requires multi-year historical model data from an unchanged or static model and observations for training. Operational models currently go through rapid changes to improve their performance. This would require long reforecasts to train a stable MOS system. Several MOS based techniques have been proposed and tested to address this drawback. For example, an updateable MOS system (Wilson and Vallée, 2002), a MOS system that relies only on the data available from past 2-4 weeks (Mao et al., 1999), CART-ACE method which combines several regression techniques such as the classification and regression trees and the alternative conditional expectation, accounting for non-linearity in bias (Gel, 2007), postprocessing method including Reanalysis (Marzban et al., 2006).

Another class of offline schemes based on the running-mean correction, requiring shorter training periods, have been successfully tested, sometimes outperforming MOS. Several of these approaches, like running the best easy systematic estimator (Woodcock and Engel, 2005) and the bias-corrected ensemble (Stensrud and Yussouf, 2003, 2005) correct systematic errors by estimating the bias at/near the observation site. These methodologies have been extended to estimate bias at grid points by using the observation operator (Hacker and Rife, 2007) and interpolating (Fan and van den Dool, 2011; Yussouf and Stensrud, 2006). These studies show that bias can be estimated using a moving training window of length as small as 7-12 days in observation dense regions.

Other offline schemes include approaches based on Kalman filter, that account for systematic as well as random errors, (Cheng et al., 2007; Delle Monache et al.,

2011), decaying average method (Cui et al., 2012). Nonlinear statistical postprocessing methods that include neural networkbased methods (Marzban, 2003) have also been investigated.

One of the major drawback of these offline schemes is that they do not impact the non-linear interaction of errors from different sources. This leads to exponential error growth till the end of forecast cycle which obscures the physical origin of the model errors and makes it more difficult to find error sources. Additionally, offline schemes require data at different lead times to correct for the corresponding forecast.

1.3.2 Online schemes: Brief Review

The online schemes, where the error correction is added as forcing to the model tendency equations, on the other hand, correct for systematic error at each model integration step. This would lead to reduction in non-linear error growth. Also, once the bias correction forcing is calculated, the scheme provides corrected forecasts at all lead times. Several such schemes that apply tendency bias corrections have been proposed and tested successfully with models ranging from simple global circulation models to more complex state-of-art coupled land-atmosphere models.

One of the earliest online correction techniques was proposed by Leith (1978). His scheme accounted for the model bias and systematic errors linearly dependent on state anomalies in the model tendency equation. He estimated the model bias correction term as the mean difference of the uncorrected model tendency and model tendency based on a reference time series. The state dependent error correction term

was estimated by minimizing the least square errors. This method also preserved the instantaneous rate of change of second moments. However, this scheme was subjected to sampling errors and requires large datasets.

Faller and Schemm (1977) applied a similar technique, STAT, to coarse-grid and fine grid versions of a modified Burgers' equation. They applied statistical corrections to correct model error arising from the coarse-grid model's inability to representing sub-grid scale processes of fine-grid models. These corrections were estimated using different regression equations at each grid point and improved the forecast skill. However, this procedure, like Leith's (1978), was subjected to sampling errors. Additionally, STAT required availability of verification data at each time step. Schemm et al. (1981) addressed this issue by introducing two new methods: (1) using time interpolated verification data, which was not successful because interpolation introduced errors and (2) applying statistical corrections less frequently, only when verification data is available. They called this method MUST. They found MUST procedure though successful in reducing artificial errors, failed when applied to (National Meteorological Center) NMC's barotropic-mesh model. They indicated that applying MUST to each grid point outperformed every other statistical scheme they tested. They tested this approach in Schemm and Faller (1986) and found that it reduced the mean square errors of the 12-hr forecasts. However, the 24-hr forecasts of certain fields, contained larger errors than the control run. The large-scale errors grew due to randomization of the residual errors by the regression equation. They suggested high pass filtering of corrections field before using it in the tendency equation.

The earliest attempts at correcting state-independent systematic errors online in operational models at NCEP were made by Saha (1992) using nudging methods. Johansson and Saha (1989) in earlier experiments with nudging compared online and offline correction methods using a simple barotropic model. They found that both methods were able to reduce the systematic errors. The online method also reduced the random errors significantly. Saha (1992) tested the same method with R30 version of the Center (NMC) T80 operational global medium range forecast model. She found that nudging, using the analysis from same dynamical model, was as successful as the then operational statistical correction, applied *a posteriori* (Alpert and Saha, 1989) in correcting systematic errors. However, with this more realistic model, the impact on random errors was only marginal and did not always lead to a reduction in errors. Again, the approach was prone to sampling errors.

DelSole and Hou (1999), tested an online approach based on Leith (1978) with a two-layer QG model. They used finite difference approximation to estimate the model tendency from the analysis and forecast while accounting for analysis errors in the modified Leith operator. They found that simply correcting the state-independent model error did not improve the forecast skill. However, correcting state dependent errors using modified Leith operator extended the forecast skill up to the limits imposed by observation error. It should be noted that the model error in this case was dominated by the state-dependent term with the bias contribution being less than 10%.

DelSole et al. (2008) tested different strategies to correct model errors in the Center for Ocean-Land-Atmosphere land-atmosphere model, version 3.2 using the

Reanalysis product by NCEP-NCAR (Kalnay et al., 1996). They found that the "nudging based on tendency errors" technique, which computed the tendency errors using slope of the least squares line fit between forecast errors at 6,12,18 and 24-hr forecasts and lead time, outperformed other techniques in term of bias reduction. This, however, did not led to improvement in the random errors, just like Saha (1992). They also emphasized that tendency errors estimated from 10 24-hr forecast were as effective as the estimates from 10 years. They further tested this methodology with the GFS (Yang et al., 2008). They estimated the tendency error by fitting the tendency errors at every 6 hr for a 21-month period to a weighted sum of a constant term plus a sine and cosine with an annual period. Using this method, they were able to reduce the temperature bias significantly, more than the offline method tested, but the impact on winds was only marginal. However, the random errors increased compared to the uncorrected GFS forecasts. They attributed this increase as being an artifact of the change in total variance of the two forecasts. Based on these results from GFS and the coupled atmosphere-ocean model, they concluded that online bias correction improves the random errors only if the bias is large to begin with.

Danforth et al. (2007) (DKM07 hereafter), developed a procedure to empirically estimate and correct the bias, diurnal errors and state-dependent systematic errors in a simplified GCM, SPEEDY (Simplified Parameterizations, primitivE-Equation DYnamics) (Molteni, 2003). They estimated the average bias of the 6-hr forecast, unlike DelSole et al. (2008) who used 24-hr forecast, by using the NCEP-NCAR Reanalysis (NNR) as an approximation of the true atmosphere, computing 6-

hr forecasts initialized from the NNR, and averaging the difference between the new NNR analysis and the 6hr forecast. Their approach is reviewed in detail in Chapter 3. Using this bias estimation, they (DKM07, Danforth and Kalnay (2008a,b) showed that: (1) It is possible to correct the model bias online by adding the (bias/6hr) term to the time derivative of each model variable; (2) online bias correction after five days was actually slightly better than the standard operational statistical bias correction made *a posteriori* from many 5-day forecasts verified against analyses; (3) Most importantly, the random errors were also significantly reduced, in contrast with DelSole et al. (2008) and Saha (1992), when the errors were corrected online rather than *a posteriori*, indicating the importance of reducing the nonlinear growth of errors; (4) The systematic errors (and corrections) in the diurnal cycle were obtained from the leading EOFs. They argued that online correction not only reduces the forecast bias but also the random forecast errors since the nonlinear model error growth due to the presence of the bias is also decreased. Following this, Li et al. (2009) found that the method gave significantly better results in model space, rather than in observation space as originally proposed by Dee (2005a).

More recently Xue et al. (Xue et al., 2015a,b) developed an online correction approach based on inverse problem to correct systematic errors in the GFS of Global and Regional Assimilation and Prediction System (GRAPES). This GFS model is different from the operational GFSv2014 used for this study and detailed later. They estimated the model tendency errors by iteratively calculating the deviation of model from NCEP reanalysis over a short time (6-hr), adding that to the model equations and then running the model again. They then used a past 30 day mean of

this tendency error with a linearly decaying weight to correct constant component of GFS systematic error. They found that this method improved forecast skill and reduced RMSE and bias for short term forecast (up to 5 days). However, this approach required running the model 80 times per day as forecasts were made every 6 hours and iterations were done 20 times to estimate model error. Thus, making this approach computationally too expensive to be adopted operationally.

1.3.3 Challenges of operational application of Online Correction Scheme

Although, it has been shown that in general, online schemes perform better than offline approaches (Section 1.3.2), the online approaches have certain limitations. Firstly, as the online correction involves adding sink/sources of heat and momentum, a large forcing may pose technical challenges (Saha, 1992) by making the model unstable, e.g: when DelSole et al. (2008) used the daily tendencies to correct land-atmosphere coupled model, the online corrected model became unstable. Secondly, the impact of online correction on the random errors has been inconsistent across the literature. Studies with simpler models with large bias showed that correcting the systematic errors also reduced the random error. However, with more complex models that was not the case. In fact, Yang et al. (2008) experiments with GFS showed an increase in the random errors. But with different training data, training lengths, definition of tendency errors, error characteristics, and frequency of sampling it is hard to make an appropriate comparison between the different experiments. Thirdly, the methodologies tested so far with GFS present a good

analysis for pedagogical use. But they can not be implemented online operationally. Yang et al. (2008) tested a methodology that requires past 21-month data, during which model changes several times and Xue et al. (2015a,b) methodology required running model 20 times every 6 hours rendering it computationally too expensive.

1.4 Objective of this dissertation

The main goal of this research is to implement and test an online approach to estimate the GFS systematic model error using 6-hr forecasts and analysis and correct for these errors online in an operational setting. More specifically this research tries to answer the following scientific questions:

1. Can a stable online correction scheme that is appropriate for operational use be designed?

2. Can short term model error be used to represent model tendency errors?

3. What are the general characteristics of model error and error growth in GFS?

4. Can an online correction aimed at correcting systematic error also impact the random component of model error?

5. How does the performance of such online systematic error correction compare with the offline methods?

We emphasize that the ultimate goal of this research is not just to empirically correct the model bias and improve the forecasts. This approach can also be used to help guide and optimize the design of subgrid-scale physical parameterizations,

more accurate discretizations of the model dynamics, boundary conditions, radiative transfer codes, and other potential model improvements that can then replace the empirical correction scheme. Though the method implemented here does not determine the exact source of the error, it provides a starting point to investigate possible causes of the error, and test the relative success of new parameterizations.

1.5 Outline of dissertation

The dissertation is structured as follows. Chapter 1 introduces the general problem and provides a brief review of the key correction schemes used to correct NWP model. Chapter 2 mainly deals with the estimation of GFS systematic errors. It first defines different types of systematic errors, then states the challenges in estimating systematic model errors and presents a brief review of past studies. Next, it presents a theoretical framework for estimating model biases using analysis as truth. The estimates of the GFS time mean and periodic systematic errors are then provided. The adaptive online correction scheme is introduced after reviewing DKM07 methodology in Chapter 3. The results after the application of the online correction are presented along with a comparison with an offline correction scheme. Finally, the conclusions and future directions are summarized in Chapter 4.

Chapter 2: Estimation of Systematic Errors in the GFS

2.1 Overview

Systematic forecast errors (SFEs) form a significant portion of the total forecast error in weather prediction models, such as the Global Forecast System (GFS), as shown in Section 1.2. Figure 1.5 shows that after 2 weeks, the range of GFS RMS temperature systematic errors reaches about one third of the total temperature forecast error range. Many studies attribute SFEs to specific deficiencies in numerical discretization of the equations of motion, parameterizations of sub-grid scale processes, or boundary conditions (e.g., Jung and Tompkins (2003a)) which lead to model bias. These errors are initially small, but as the model is integrated in time the errors grow and interact nonlinearly with other systematic and random errors until the model loses all forecast skill. This chapter aims at estimating the GFS model deficiencies that leads to Systematic Forecast Errors (SFEs) in the period 2012 - 2016 using Analysis Increments (AIs). The results of this chapter have been published in Bhargava et al. (2018).

2.1.1 Characterizing model error

Model errors can be classified into random errors, whose time average is zero, and systematic errors (Dalcher and Kalnay, 1987; Murphy, 1988), described in eqn. (1.1). The SFEs include mean and periodic forecast biases, the latter including those associated with the annual cycle and the diurnal cycle, and also state-dependent errors that are highly correlated with certain states of atmosphere, for example, associated with the presence of short or long-term anomalies, such as the phase of El Niño (DKM07). According to the definition in eqn. (1.1), periodic and state-dependent errors would be also considered as random. To address this problem, here we define the model errors as in DKM07,

$$\mathbf{x}_m^e = b + \sum_{l=1}^{L} \beta_l(t) e_l + \sum_{n=1}^{N} \gamma_n(t) f_n + \mathbf{x}_m^{rand}, \tag{2.1}$$

where the model error, $\mathbf{x_m^e}$, consists of 1) a time mean bias term, b, obtained by averaging the forecast errors b=($\bar{x^e}$) over a certain training period, 2) periodic errors estimated using leading EOFs e_l from the anomalous error field diurnal cycle, 3) the state-dependent model error component given by the leading Singular Value Decomposition (SVD) modes f_n of the covariance between the coupled model state anomalies and the corresponding error anomalies. The last term in eqn. (2.1) represents non-periodic random errors. L and N are the number of retained leading modes of EOFs and SVDs, respectively. The collective contribution of the periodic, state-dependent and random errors as defined in eqn. (1.1) is referred to as the apparent random error or the unbiased component of error in the text.

2.1.2 Chapter outline

In this chapter, we first develop a theoretical framework for estimating model biases using AIs. Then using this, mean and periodic model biases of the NCEP operational GFS/GDAS system, for the period 2012-2016, are estimated from the 6-hr analysis increments (AIs), before the forecast errors grow nonlinearly using a methodology similar to DKM07 (explained in section 2.2.2). AIs are the difference between the gridded analysis and forecast, with the former providing our best gridded estimate of the true state of the atmosphere in absence of a true atmospheric state. In addition to suggesting causes for these model errors and exploring the impact of changes in the GFS system, we evaluate their potential use as input to an empirical online correction scheme as introduced by Danforth and Kalnay (Danforth and Kalnay, 2008a,b; Danforth et al., 2007).

The chapter is organized as follows. Firstly, the GFS model is described in the section 2.2. Section 2.3 reviews the key past studies. Section 2.4 describes the methodology used to estimate model systematic errors in GFS by first describing the challenges in estimation of model error in section 2.4.1. Section 2.4.2 then presents a theoretical framework to estimate model biases using analysis as verification along with listing the assumptions and limitation involved. In section 2.5, we examine the structure and evolution of the biases. We also compare the bias correction for the 2012-2014 period, during which few model changes took place, to the final 2015-2016 period when major model changes and changes to the SST boundary condition took place. In section 2.6, sub-monthly periodic biases are estimated and represented

using EOFs to provide evidence that a low dimensional approach can also be used to correct the dominant diurnal and semi-diurnal errors. A summary and discussion of the results and details regarding the online correction experiments to be performed with the estimated systematic and daily errors is presented in section 2.7.

2.2 Model details: GFS

The GFS is a three-dimensional hydrostatic global spectral model with current operational resolution of T1534 from 0-10 days and T574 from 10-16 days. The model uses 64 hybrid sigma-pressure levels (Sela, 2009) in the vertical, defined as: $p(x, y, t) = \sigma_1 * p_s + \sigma_2$ so that they become parallel pressure levels at high altitudes, σ_1 and σ_2 are given parameters, and p_s is the surface pressure. Here, results at 7 representative model levels, including the bottom two levels, the top level, and four model levels in between are presented (table 2.1). The GFS is run four times a day and forecasts are made available every hour for the first 120 hours, then every 3 hours for up to 10 days and then every 12 hours. The GFS analysis is run twice per cycle: the "early" GFS run that provides 16-day forecasts, and the "final" GDAS (Global Data Assimilation System) run that assimilates late-arriving observations. The 'GDAS run' also includes running a short forecast (nine hours), initialized with GDAS analysis, to provide the first guess to both the gfs and gdas for the following cycle. The GDAS currently uses a hybrid four-dimensional ensemble variational formulation (Buehner et al., 2013; Kleist and Ide, 2015).

The GFS/GDAS system is updated regularly to improve its performance. At

Table 2.1: Model levels shown and their parameters

Model level	Parameter σ_1	Parameter σ_2	Pressure assuming p_s =1000mb
1	1	0	1000
2	0.995	0	995
7	0.954	116.899	950
14	0.827	2051.15	850
25	0.393	12344.49	500
35	0.506	15683.489	200
64	0	64.27	0

the beginning of our study period, January 2012, the GDAS was based on the 3DVar Gridpoint Statistical Interpolation (Wu et al., 2002). It used T574 resolution semi-implicit Eulerian discretization, with the lower SST boundary condition over the oceans provided by the weekly averaged Optimal Interpolation SST (Reynolds and Smith, 1994). Beginning in May 2012, a hybrid 3DVar-ENKF data assimilation system (Wang et al., 2013), which makes use of a background error estimate from a combination of a lower resolution Ensemble Kalman Filter and a static background error, replaced the prior 3DVar. In January 2015, GFS transitioned to a two time-level T1534 semi-implicit semi-Lagrangian discretization and switched to the high resolution daily real time global (RTG) SST product (Thiébaux et al., 2003). In May 2016, the hybrid data assimilation system was upgraded to the current operational 4D hybrid ensemble-variational data assimilation system

(Buehner et al., 2013; Kleist and Ide, 2015). Modifications to the Advanced Microwave Sounding Unit radiance assimilation that allow for more "all-sky" data to be assimilated (Zhu et al., 2016) were also added. The latest update in 2017, implemented Near Surface Sea Temperature (NSST) to replace RTG SST to provide more realistic ocean boundary conditions. NSST gives the vertical profile of sea temperatures near surface that include the impacts of diurnal warming and sublayer cooling physics processes. The details of the evolution of GFS are described at: www.emc.ncep.noaa.gov/gmb/STATS/html/model_changes.html.

2.3 Key Literature

The most common way to estimate state independent model errors is to calculate the difference between temporal mean of the model short term forecast, say 1-day or less, and a verifying truth (DelSole et al., 2008; Saha, 1992), over some training period ranging from 10 days to several years. Short term forecasts have the advantage of identifying the local source of problems. The forecast length should be short enough such that the forecast errors are still growing linearly and errors from observations and/or analysis have not started interacting nonlinearly yet.

Klinker and Sardeshmukh (1992) took this method of model error estimation using short term forecast to its limit by investigating the initial tendency errors in operational European Center for Medium-Range Forecast (ECMWF) model to identify the source of errors. They pointed out that though it is difficult to separate model errors from forecast errors, initial tendency errors are calculated close to

the best estimate of truth. Hence, they should be associated with model tendencies than the inaccuracies in the initial conditions. Also, studying initial tendency errors, when the error in the model's degree of freedom is decoupled, can provide insights into origins of model error. They inferred that the major source of model error was model's gravity wave parameterization by comparing the tendency error and tendency from individual model components.

Leith (1978) formulated a methodology to estimate state-dependent systematic error along with the bias. For a given model, $\dot{x} = \mathcal{M}(x)$, Leith sought an improved model of the form, $\dot{x} = \mathcal{M}(x) + Lx + +b$, where Lx provides the state-dependent estimate of the model error and b is the state-independent component of systematic error, also referred to as bias. He estimated b and L by minimizing the mean square tendency error of the improved model. However, direct computation of Lx is computationally prohibitive.

Based on Leiths methodology, DKM07 developed an approach to estimate the bias, periodic component and the state-dependent systematic error using SPEEDY model. They calculated the bias by taking the mean of differences between the 6-hr model forecast and the NCEP-NCAR reanalysis NNR over a month. They found that the periodic error was dominated by the errors in diurnal cycle and could be represented using just 2 leading EOF modes. This was expected as SPEEDY lacks diurnal forcing. To estimate the state-dependent component, they developed a low dimensional approach based on the singular value decomposition (SVD) of the anomalous error and state covariance. They showed that accounting for these estimates in the model tendency equations improved not only the systematic but

also the random component of error in SPEEDY forecast.

Recent studies have also shown that short term forecast errors can be used as an estimate of model errors when initial condition errors are small to begin with even in the presence of observation/analysis errors.

Orrell et al. (2001) developed a mathematical framework to investigate the error growth due to inaccuracy in initial condition and model error. They demonstrated that errors in the initial conditions are propagated by the tangent linear model and the local model drift caused by model errors at any time is given by summation of short-term deviations of forecast from truth during that period. They also mentioned that in cases where the analysis error is negligible, model error is equal to short term forecast error. They used this technique to study the errors in ECMWF using analysis as the truth. They estimated the error due to model by summing the 6-hr AIs for up to 1 day and analysis minus 24-hr forecasts at longer lead times. They found that model errors dominate the forecast error for about 3 days.

Privé and Errico (2013) investigated the role of initial condition and model error with an OSSE using Global Earth Observing System version 5.7.1 and the Gridpoint Statistical Interpolation (GSI, Kleist et al. (2009)) data assimilation system. They showed that model error not only impacted forecast skill but also the quality of analysis which retained significant portion of the systematic model error. They also found that the even though the observation errors degraded the medium range forecast skills for perfect model, they had much less impact on forecast skill when model errors were present.

2.4 Methodology: Estimating GFS Systematic Model error from 6-hr analysis Increments

2.4.1 Challenges

The first challenge and the first step towards estimating systematic model errors is the choice of verifying data. As mentioned in section 2.2.1, the options available are the bias corrected observation, native analysis or reanalysis. Most online correction schemes use either a reanalysis product (Danforth and Kalnay, 2008a,b; Danforth et al., 2007; DelSole et al., 2009, 2008) or native analysis (Saha, 1992; Xue et al., 2015a; Yang et al., 2008). The primary reason that online correction schemes avoid using observations is that observations are not present at the grid points or at the time required and thus have the problem of spatial and temporal interpolation which might introduce errors (Schemm et al., 1981; Xue et al., 2015b) and not be as accurate a choice for truth. Schemm et al. (1981) tested using verification data available less frequently and found that using time interpolation techniques was not successful in reducing model errors as interpolation introduced errors.

One way to avoid these issues is to use the reanalyses as the verifying truth and the initial conditions. Reanalyses, which use a different model, introduces large initial errors during spin up (DelSole et al., 2008; Klocke and Rodwell, 2014). However, the reanalysis uses a different dynamical model than the one being corrected. This inconsistency may lead to "initial shock" introducing artificial error in the forecasts (DelSole et al., 2008; Klocke and Rodwell, 2014). The interpolation of reanalysis to

model grid, if not present on the required grid, may also give rise to errors that do not truly represent forecast errors (Klocke and Rodwell, 2014). Hence, calculating errors at short lead times would not be as effective.

The last option is to use native analyses which is present at the required time and on the required grid and is generated using the same dynamical model. However, using the native analysis that has been generated using bias corrected observations also has its own limitation. Even in absence of observation bias, if the forecast is biased the resulting analysis would be biased. Privé and Errico (2015) showed using an OSSE that when verified against truth, analyses has errors similar to those in short-term forecasts. Hence, AIs underestimate the true error forecast during early forecast period. These estimates approached true error after at least 48 hours which results in an estimation of error growth that is steeper than the actual error growth. They found that highest discrepancies existed at large scales for wind and at small scales for the temperature and humidity. Despite these limitations, Privé and Errico (2015) also suggested analyses being the best estimate of truth in real world where truth is unknown. We discuss the use of native analysis as verifying truth and associated assumptions and limitations in the next section.

The second and more difficult challenge is to separate model error from forecast error. Forecast errors in a model can arise from inaccurate initial conditions and/or model deficiencies, as mentioned earlier. Initial condition errors arise not only due to errors in observation but also depend on the observation network and how the DA system handles background and observations. Though attempts have been made to improve both (refer Chapter 1 for details), it is difficult to separate

their respective contribution to the total forecast error (Dee, 2005b; Lorente-Plazas and Hacker, 2017). Comparison of different models and sensitivity to changing specific parameters or running Observation System Simulation Experiments (OSSE) (Privé and Errico, 2013, 2015) do provide some insights into the model error and origins but fail to quantify the model deficiencies in absolute terms. With different models, parameters, and data assimilation schemes, sharing similar assumptions, comparing models just provides a lower bound of the model error. Another class of methods used to separate these two are based on analyzing error growth. This involves assuming exponential error growth of initial condition errors and linear growth of model errors (Leith, 1978), additionally including a saturation term for initial condition error growth (Dalcher and Kalnay, 1987). Kalman filter based methods (Lorente-Plazas and Hacker, 2017; Pauwels et al., 2013) have also been used to estimate the observation and forecast biases. These methods have only been tested in an OSSE framework where the structure of model and observation error is known. Also, a major assumption in these techniques is that the model state and the systematic errors are uncorrelated. Hence, the state-dependent errors are ignored. To address this problem, we here assume that observations have been bias corrected before and no other observation related biases are present and hence do not contribute to biases in forecast error.

2.4.2 Model bias estimation: Theoretical framework

In this section, we provide a theoretical framework for estimating model bias using native analysis as verifying truth. We follow the approach of Klinker and Sardeshmukh (1992) and estimate the time independent bias by calculating the tendency bias. Let's assume that the evolution of true atmospheric state $\mathbf{x_t}$, projected in model space, is given by

$$\frac{\mathbf{dx}_t}{dt} = \mathcal{T}(\mathbf{x_t}(\mathbf{t})) \tag{2.2}$$

and the model state $\mathbf{x_b}$ is given by the model tendency equation as below,

$$\frac{\mathbf{dx_b}}{dt} = \mathcal{M}(\mathbf{x_b}(t)) \tag{2.3}$$

The model tendency bias, α_m, where the subscript 'm' stands for model, for small dt is given by,

$$\alpha_m = \langle \mathcal{M}(\mathbf{x_t}(\mathbf{t})) - \mathcal{T}(\mathbf{x_t}(\mathbf{t})) \rangle \tag{2.4}$$

here $\langle . \rangle$ denotes mean over a large number of forecasts. The equations (2.4) can be discretized and written with limits on the integration time Δt as

$$\alpha_m = -\left\langle \lim_{\Delta t \to 0} \frac{\mathbf{x_t}(t+\Delta t) - \mathbf{x_b}(t+\Delta t)}{\Delta t} \right\rangle \tag{2.5}$$

The model state $\mathbf{x_b}$ is calculated by integrating the eqn. (2.3) from the true atmosphere state as, $\mathbf{x_b}(t+\Delta t) = \int_t^{t+\Delta t} \mathcal{M}(\mathbf{x_b}(s))ds$ where $\mathbf{x_b}(s) = \mathbf{x_t}(s)$ when $s = t$ and $\mathbf{x_t}(t+\Delta t)$ is the true state at $t+\Delta t$. Thus, when both the models, eqns. (2.2) and (2.3), are integrated from the true state for a small integration time Δt, the model bias, $\mathbf{b_m} = \alpha_m * \Delta t$.

In realistic systems, the true state in unknown. The most commonly used proxy for the truth is the analysis. Analysis is estimated by combining short-term forecast/background and observations using a data assimilation scheme. Using analysis as verifying truth on the right hand-side of eqn. (2.5)

$$\begin{aligned} & -\left\langle \lim_{\Delta t\to 0} \frac{\mathbf{x_a}(t+\Delta t) - \mathbf{x_b}(t+\Delta t)}{\Delta t} \right\rangle \\ = & -\left\langle \lim_{\Delta t\to 0} \frac{\mathbf{x_t}(t+\Delta t) + \mathbf{e_a}(t+\Delta t) - \mathbf{x_b}(t+\Delta t)}{\Delta t} \right\rangle \\ = & -\left\langle \lim_{\Delta t\to 0} \frac{\mathbf{x_t}(t+\Delta t) - \mathbf{x_b}(t+\Delta t)}{\Delta t} + \frac{\mathbf{e_a}(t+\Delta t)}{\Delta t} \right\rangle \end{aligned} \tag{2.6}$$

where the analysis error is given by $\mathbf{e_a}(t+\Delta t) = \mathbf{x_a}(t+\Delta t) - \mathbf{x_t}(t+\Delta t)$. The eqn. (2.6) provides the true measure of tendency bias if analysis error, the term highlighted in red, is zero or random. In deriving this equation we used analysis as the verifying truth but assumed that the model (eqn. 2.3) is initialized from the true state. However, as the true state is unknown, the model has to be initialized using analysis, native or non-native. This would further introduce errors in the forecast. As, we have defined $\mathbf{x_b}$ as model integrated from true state, we use the notation $\mathbf{x_{b,anl}}$ to represent model integrated from the specified analysis state. So, if we use the analysis as the verifying truth as well as the initial condition for the model, it becomes

$$\begin{aligned} & -\left\langle \lim_{\Delta t\to 0} \frac{\mathbf{x_a}(t+\Delta t) - \mathbf{x_{b,anl}}(t+\Delta t)}{\Delta t} \right\rangle \\ = & -\left\langle \lim_{\Delta t\to 0} \frac{\mathbf{x_t}(t+\Delta t) + \mathbf{e_a}(t+\Delta t) - (\mathbf{x_b}(t+\Delta t) + \mathbf{e_{m,anl}}(t+\Delta t))}{\Delta t} \right\rangle \\ = & -\left\langle \lim_{\Delta t\to 0} \frac{\mathbf{x_t}(t+\Delta t) - \mathbf{x_b}(t+\Delta t)}{\Delta t} + \frac{\mathbf{e_a}(t+\Delta t) - \mathbf{e_{m,anl}}(t+\Delta t)}{\Delta t} \right\rangle \end{aligned}$$

(2.7)

where $\mathbf{e_{m,anl}}(t+\Delta t)$ represents the additional model error, apart from the model bias, that arises from providing inaccurate initial conditions. Using eqns. (2.7) and (2.5), we get

$$-\left\langle \lim_{\Delta t\to 0} \frac{\mathbf{x_a}(t+\Delta t)-\mathbf{x_{b,anl}}(t+\Delta t)}{\Delta t} \right\rangle = \alpha_m - \left\langle \lim_{\Delta t\to 0} \frac{\mathbf{e_a}(t+\Delta t)-\mathbf{e_{m,anl}}(t+\Delta t)}{\Delta t} \right\rangle \quad (2.8)$$

In other words, for a very small time step, the negative average analysis increment divided by the time step provides the tendency bias along with the average difference between the analysis error and the model error due to inaccurate initial conditions. Let's denote the additional term in red in the equation above using β.

$$\beta = -\left\langle \lim_{\Delta t\to 0} \frac{\mathbf{e_a}(t+\Delta t)-\mathbf{e_{m,anl}}(t+\Delta t)}{\Delta t} \right\rangle \quad (2.9)$$

Now there are two possibilities (1) $|\beta| = 0$, (2) $|\beta| > 0$.

1. **Case I:** $|\beta| = 0$: In this case the average analysis increments provide a true measure of the model bias for a very small time step and large sample, for the following cases

   (a) The analysis error is random and when propagated by model, the resulting background error due to these imperfect initial conditions is also random. This is valid for bounded systems with infinitely large sample size. However, these additional terms should be considered when the sample size is finite.

   (b) The average analysis error and the average forecast error arising due to errors in the analysis are equal. Though, this seems to be unlikely given

the assumptions involved in the data assimilation system that generates the analysis.

2. **Case II:** $|\beta| > 0$: In this case, averaged analysis increment might not provide an accurate measure of model bias and have some bias of their own. We focus on what this non-zero term is by exploring the analysis errors more next.

For a nonlinear system with Gaussian distribution of the background state, the data assimilation equation (using the LETKF) used to generate the analysis by attempting to optimally combine the information from unbiased forecast and observation errors is given by,

$$\begin{aligned} \mathbf{x_a} &= \mathbf{x_{b,anl}} + \mathbf{K}[\mathbf{y_o} - h(\mathbf{x_{b,anl}})] \\ \mathbf{K} &= \mathbf{B}\mathbf{H^T}(\mathbf{R} + \mathbf{H^T}\mathbf{B}\mathbf{H})^{-1} \end{aligned} \tag{2.10}$$

where $\mathbf{K}$ is the Kalman gain, h is the nonlinear observation operator, $\mathbf{B}$ is the background error variance and $\mathbf{R}$ represents the observation error variance. Using these equations, the analysis error can be written as

$$\mathbf{e_a}(t + \Delta t) = \mathbf{K}\mathbf{e_o}(t + \Delta t) + [\mathbf{I} - \mathbf{K}\mathbf{H}]\mathbf{e_b}(t + \Delta t) \tag{2.11}$$

where $\mathbf{e_b}$ is the total error in the background, both due to imperfect initial conditions and model biases. The eqn. (2.11) shows that error in the analysis can arise due to:

1. **Analysis errors from observation related errors:** These can arise from

   (a) Direct observation errors that involve measurement biases and representative errors.

(b) Errors related to observation assimilation that include errors in the observation operator, highly nonlinear observation operator such that higher order nonlinear terms cannot be neglected and incorrectly assuming unbiased observation error.

2. **Analysis errors from background errors:** These arise from the errors in the background/forecast due to model deficiencies and poorly posed initial conditions, for example inaccurately assuming unbiased observation error, or presence of model bias, both would lead to bias in the initial conditions.

3. **Incorrect data assimilation assumptions:** These include incorrectly assuming Gaussian statistics for a nonlinear system (which is not always true for highly nonlinear system), ignoring higher order nonlinear terms, incorrectly assuming unbiased observation and model errors.

Next we analyze, the second term that contributes to β, i.e. the model error due to inaccurate initial conditions. These can be written as

$$\mathbf{e_{m,anl}}(t+\Delta t) = M(t, t+\Delta t)\mathbf{e_a}(t) + \epsilon(t, t+\Delta t) \tag{2.12}$$

where M is the tangent linear model and ϵ represents higher order terms that arise due to nonlinearity. Using eqns. (2.11) and (2.12), the eqn. (2.9) becomes

$$\begin{aligned}
\beta = & -\left\langle \lim_{\Delta t\to 0} \frac{\mathbf{K e_o}(t+\Delta t) + [\mathbf{I}-\mathbf{KH}]\mathbf{e_b}(t+\Delta t) - M(t,t+\Delta t)\mathbf{e_a}(t) - \epsilon(t,t+\Delta t)}{\Delta t} \right\rangle \\
= & -\left\langle \lim_{\Delta t\to 0} \frac{\mathbf{K e_o}(t+\Delta t) + [\mathbf{I}-\mathbf{KH}] * [\alpha_m * \Delta t + M(t,t+\Delta t)\mathbf{e_a}(t) + \epsilon(t,t+\Delta t)]}{\Delta t} \right\rangle \\
& -\left\langle \lim_{\Delta t\to 0} \frac{M(t,t+\Delta t)\mathbf{e_a}(t) + \epsilon(t,t+\Delta t)}{\Delta t} \right\rangle
\end{aligned}$$

$$
\begin{aligned}
= \quad & -\left\langle \lim_{\Delta t \to 0} \frac{\mathbf{K}\mathbf{e_o}(t+\Delta t)}{\Delta t} \right\rangle - \left\langle \lim_{\Delta t \to 0} \frac{[\mathbf{I}-\mathbf{KH}] * [\alpha_m * \Delta t]}{\Delta t} \right\rangle \\
& + \left\langle \lim_{\Delta t \to 0} \frac{\mathbf{KH} * [M(t, t+\Delta t)\mathbf{e_a}(t) + \epsilon(t, t+\Delta t)]}{\Delta t} \right\rangle
\end{aligned}
\tag{2.13}
$$

The right hand side of the eqn. (2.13) is non-negligible when

1. Large observation related errors and biases are present. This makes the first and the third terms on the right hand-side of eqn. (2.13) significant. Though it's likely that these terms might partly balance each other.

2. Large model bias and small magnitude of **KH** is present. This would lead to underestimation of the model tendency bias by averaged analysis increments.

3. There's amplification of initial condition by the tangent linear model and significant nonlinear terms in the model error.

A clear decomposition and detailed analysis and quantification of these term that form β is out of the scope of this work. The analysis gets more complex as we begin to account for error evolution due to cycling the data assimilation system and model runs. However, we try to simplify this for a special case where the observations are unbiased, observation and background errors have Gaussian statistics with insignificant high order nonlinear terms is very small, such that $\mathbf{KH} \approx \mathbf{I}$. This reduces the eqn. (2.13) to

$$
\begin{aligned}
\beta = \quad & -\left\langle \lim_{\Delta t \to 0} \frac{\mathbf{KHM}(t, t+\Delta t)\mathbf{e_a}(t)}{\Delta t} \right\rangle \\
= \quad & -\left\langle \lim_{\Delta t \to 0} \frac{M(t, t+\Delta t) * (\mathbf{K}\mathbf{e_o} + (\mathbf{I}-\mathbf{KH})\mathbf{e_b}(t)) + \epsilon(t, t+\Delta t)}{\Delta t} \right\rangle \textbf{using eqn. 2.11}
\end{aligned}
$$

$$= \quad -\left\langle \lim_{\Delta t \to 0} \frac{M(t, t+\Delta t) * e_o}{\Delta t} \right\rangle \tag{2.14}$$

Thus, in our case average analysis increments over a large sample and small time step Δt, provide the estimate of the model bias along with an addition β term. This β term is negligible only when the observation errors propagated by the tangent linear model are random and all the assumptions stated earlier are valid.

In this chapter, we use the 6-hr analysis increments to estimate the model bias, assuming the observations are bias corrected before being assimilated. As there's cycling of the model run and data assimilation systems and several model integration steps are involved instead of $\lim \Delta t \to 0$, additional bias may itself be present in the bias estimation. Additionally, the observation bias correction may not be independent of model bias. Several observation bias correction schemes use analysis, which itself has biases present due to model bias, as the verifying truth. Accounting for how these observation bias correction schemes impact the estimation of model bias is left for future studies. Despite these limitations, in the absence of any available truth, analysis increments serve as the best first estimate of model bias for operational systems.

2.4.3 Application to GFS

To estimate the model bias, we take advantage of the GDAS, which attempts to optimally combines the 6-hr forecast, or background, with the all available new high-quality bias corrected observations, creating a new analysis. The analysis is

the best estimate of truth we have after combining the model forecasts and the observations. We use the time average of the AIs as the model bias correction over 6-hr, the negative of which is the 6-hr model bias, assuming bias-free observations. An important advantage of this approach is that over 6-hr, the forecast error growth is linear (Jung and Tompkins, 2003a,b; Klinker and Sardeshmukh, 1992; Vannitsem and Toth, 2002; Xue et al., 2015a). Hence, the average 6-hr AIs give the best estimate of the model bias before the errors start growing nonlinearly. Here, we assume that the observations are bias corrected. In practice the observations are bias corrected either offline or online using variational bias correction methods (Dee and Uppala, 2009; Zhu et al., 2015). These methods lead to observations being relaxed to background or analysis. Both of these might introduce model bias in the observations itself. Accounting for these is a more complex problem and out of scope of this dissertation. We do suggest that this problem be addressed before any operational implementation.

We use the 6-hr analysis and forecasts for surface pressure, temperature, winds and specific humidity provided by the operational GFS. The GFS horizontal resolution was T574 until January 2015 when it transitioned to T1534. For convenience, we remap the variables, averaging all 4 daily cycles over a month, through the full period of interest 2012-2016 onto a uniform lower resolution T254 grid, to match the resolution at which we have access to the AIs. We explored the impact of reducing resolution by projecting the analysis increments on different resolution grids. We found that reduction in resolution has essentially no impact on our analysis, as illustrated in Fig. 2.1. We begin by focusing on seasonal model bias correction,

which we estimate as the seasonal average (DJF, MAM, JJA, and SON) of the AIs during the five years 2012-2016. The temporal stability of the seasonal bias is evaluated, by comparing the spatial patterns of the seasonal AIs for the first three years (2012-2014) and evaluating their similarity using anomaly correlations.

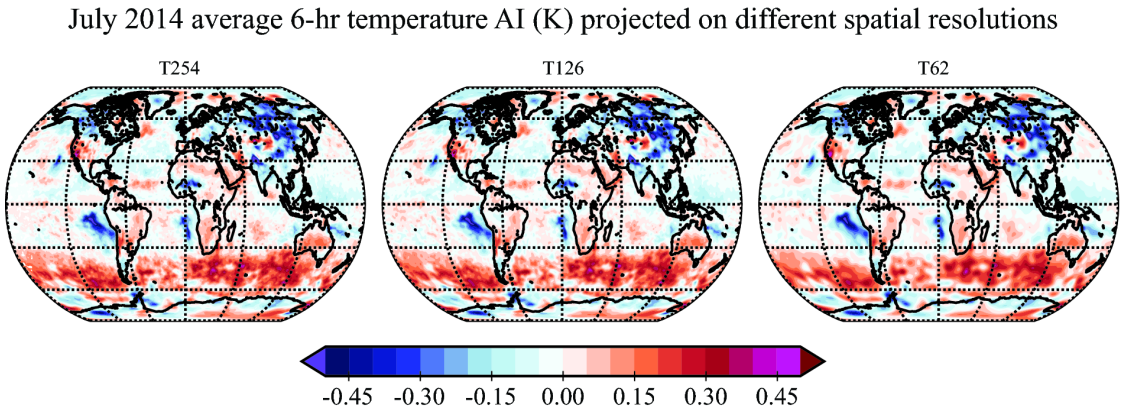

Figure 2.1: 6-hr model bias for surface temperature, averaged over all 4 cycles daily for July 2014, projected on three spatial resolutions: T254, the original resolution of data provided, (left), T126 (middle) and T62 (right). The patterns of bias remain essentially the same, indicating that the scales of the model bias are well resolved by T62.

To identify the systematic components of the periodic AIs at sub-monthly

scales we first calculate the anomalies of the 6-hourly AIs with respect to their monthly averages. We then decompose these anomalies into a complete set of 120 Empirical Orthogonal Functions (EOFs) and corresponding principal component time series. These EOFs are geographically weighted, so both the spatial patterns and the time series are orthogonal over the surface of the globe. We find that these are dominated by the diurnal and semi-diurnal components (see section 2.6).

2.5 Time mean Error: Seasonal bias

In this section, we examine the structure and evolution of the seasonal cycle biases. We begin by examining the seasonal biases and compare the bias corrections for the initial three-year period 2012-2014, during which few model changes took place, to the final 2-year period, 2015-2016 with major model and boundary condition changes.

We first explore how the global mean error in GFS forecasts changes with height (Fig. 2.5). The estimated GFS global mean RMS error of temperature and winds varies from about 0.15 K to 0.2 K and 0.17 m/s to 0.21 m/s , from the surface to level 54 (approximately 13 mb). They then grow steeply, presumably because of the effects of the artificial rigid upper boundary which introduces errors in the radiative balance and generates spurious dynamic instabilities (Hartmann et al., 1996) that remain attached to the top. Peaks in both temperature and wind RMS errors occur at about 35th model level (∼200-250 mb) with peak in winds being extremely prominent. This is associated with the large errors in the estimation of

jet stream. Specific humidity error increases from 0.05g/kg near surface to 0.1 g/kg at 850 mb, decreasing so that by 300 mb the air is dry. Here, we present results only for the surface and ∼850 mb.

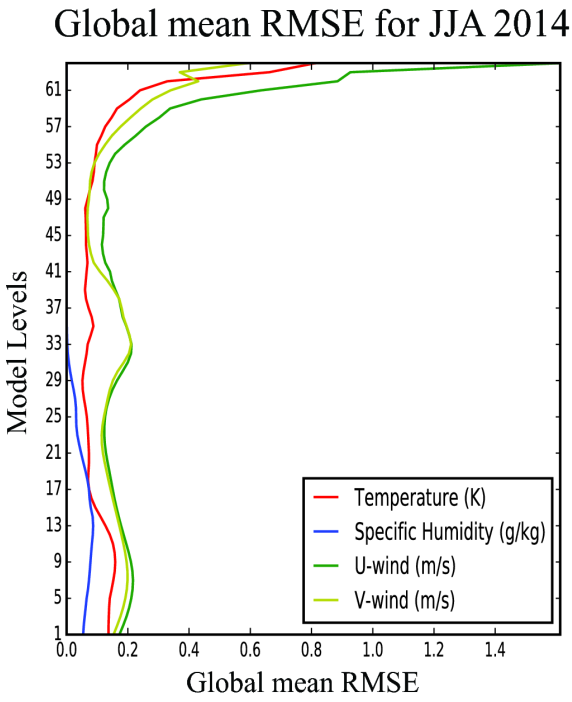

Figure 2.2: Global mean temperature, specific humidity and winds error vs model level for JJA 2014. The increase in error for levels above 53 is discussed in the text.

Despite major changes made to the data assimilation scheme in May 2012, the bias corrections retain their major features throughout 2012 to 2014 (Figs. 2.3 and 2.4). In general, the model tends to underestimate surface pressure over the land and overestimate it over the ocean, except the regions of warm pools at the Gulf of Mexico, North Atlantic, and Bay of Bengal (Fig. 2.3). The surface pressure bias over the land peaks in local summer and is lowest during local winter. Conversely, over ocean the high bias peaks during the local winter.

South and East Asia show a -10 to -20 Pa erroneously low forecast surface pressure during JJA (Fig. 2.3). This is the result of erroneously warm and dry forecast air, which peaks during the summer monsoon (Fig. 2.4). A possible explanation is that the monsoon winds carrying moisture in from the Southern Hemisphere are erroneously weak as seen in the bottom panel of fig. 2.4 that shows weaker southerly winds indicated by red color off the coast of Somalia. Near the Equator the elevated humidity associated with the ITCZ is spread too wide meridionally, with too weak convergence, so that the ITCZ itself is too dry and the Equator is too moist (Fig. 2.4).

Erroneously warm temperatures are also present over the subsidence zones west of South Africa and South America (Fig. 2.4). For many GCMS high temperatures in these locations are likely due to an inability to maintain sufficient stratus clouds (Lien et al., 2016; Zheng et al., 2011). What is different here is that the biases are more strongly concentrated in the Southern Hemisphere and are displaced a few degrees westward from the coast. These areas also have a dry bias of 0.3 to 0.6 g/kg. A cold bias is present over the oceans in higher latitudes during local summer with the bias being more prominent over the Southern Ocean. This is mainly due to under representation of clouds in the GDAS analysis which makes the GDAS analysis warmer (Yang et al., 2013). Accompanying the cold bias over the Southern Ocean is a positive surface pressure bias. Interestingly, the biases in surface pressure increased after the data assimilation changes made in May 2012. The winds in this region show a north-easterly bias. Over the Southern Ocean (60°S-40°S) surface temperature forecasts are -0.2K/6-hr erroneously cool (Fig. 2.4), while the intense

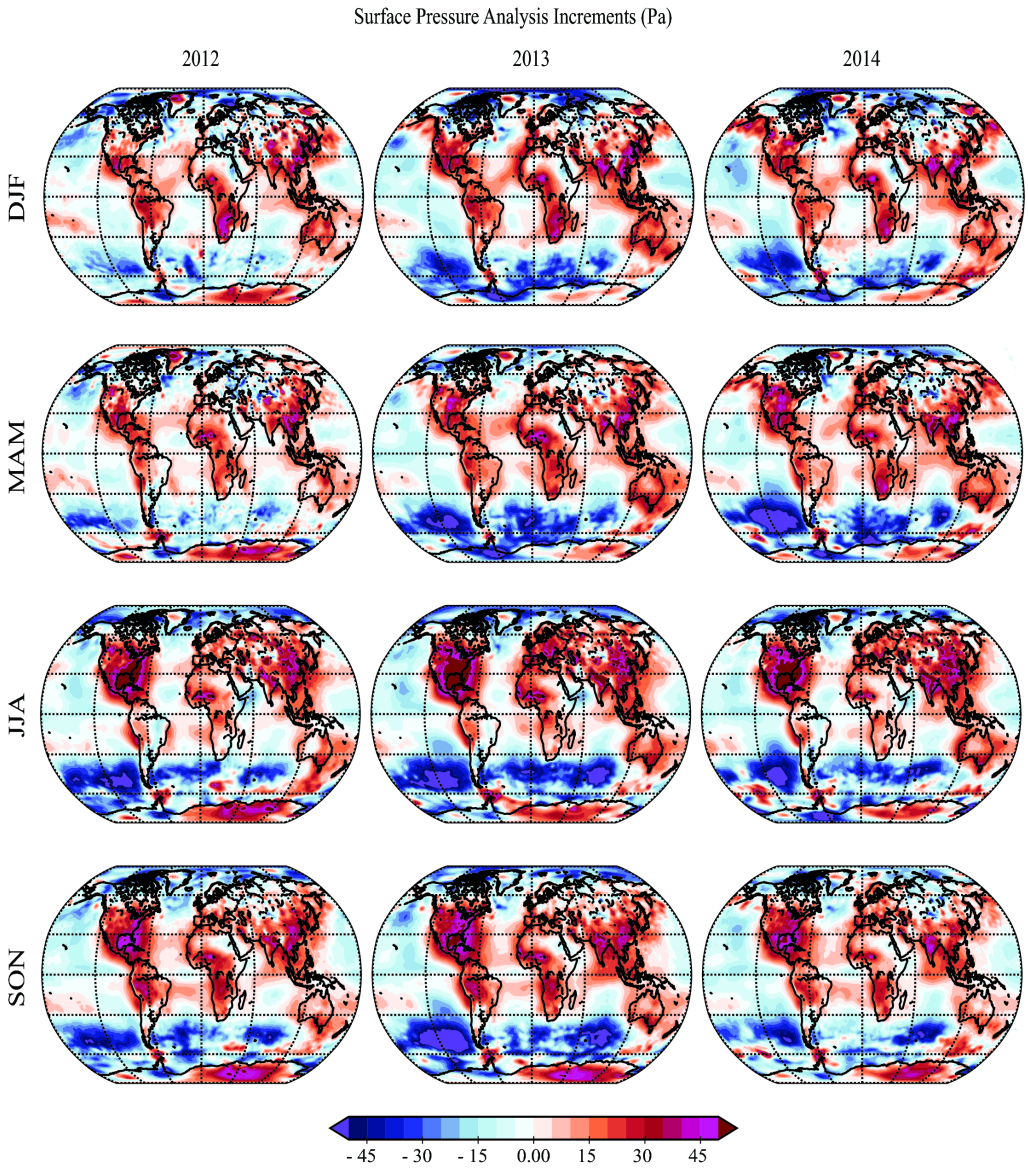

Figure 2.3: Seasonally averaged surface pressure AIs (Pa) for 2012 to 2014 (left to right). Forecast surface pressure is generally too high (cool colors) over the oceans, except near coasts, and too low (warm colors) over the continents. Seasonal mean AIs remain relatively consistent for the 3 years.

easterlies that dominate in this latitude zone are displaced 5° too far northward. This bias pattern over the ocean in the Northern Hemisphere is also found in var-

ious GCM simulations and is hypothesized to be due to inaccuracies in simulation of North Atlantic storms (Chapman et al., 2007).

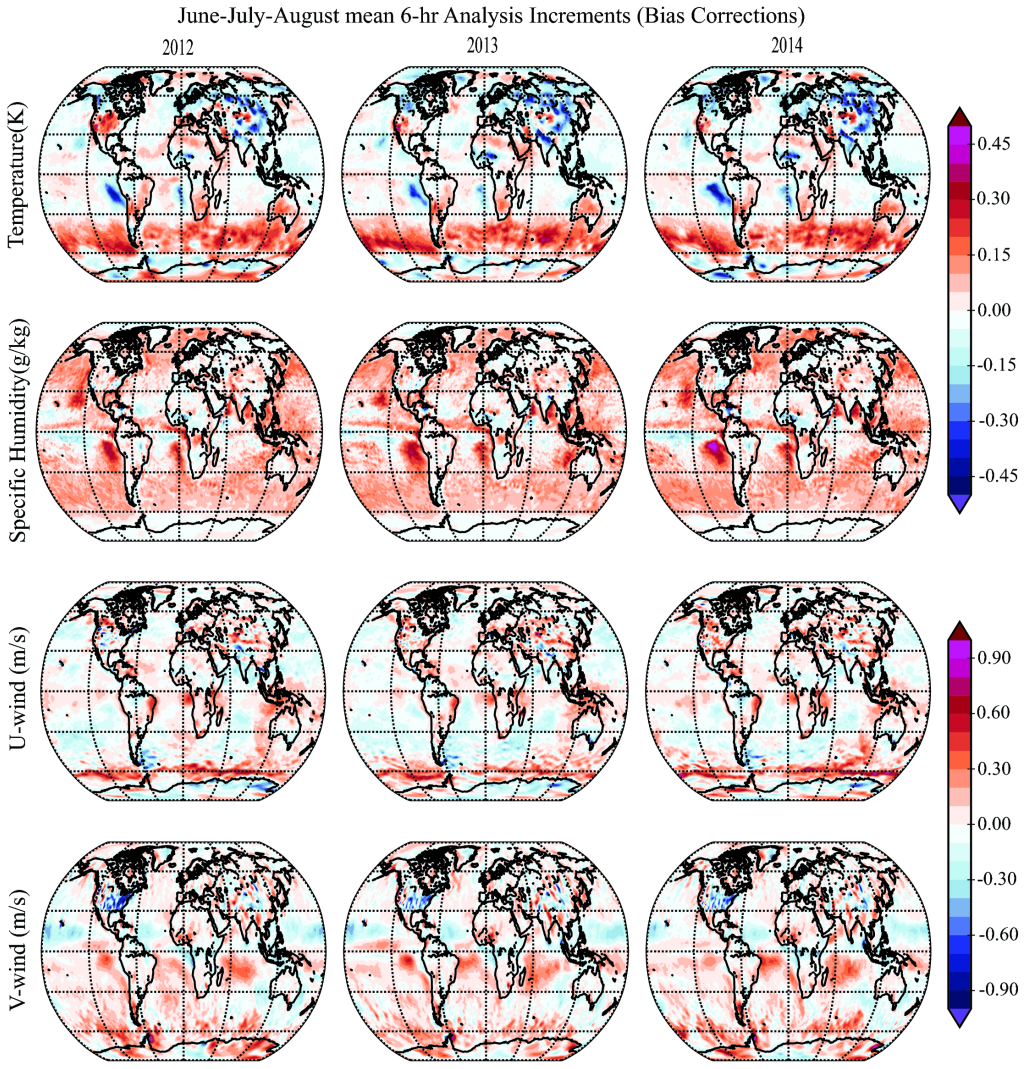

Figure 2.4: JJA averaged AIs for the years 2012 (left), 2013 (middle) and 2014 (right) at approximately 850 mb. The AIs remain quite consistent from 2012 to 2014.

We next explore how the AIs change when progressing from the years 2012-2014 to 2015-2016 (Fig. 2.5). The most striking changes occur over the oceans. There we see a reduction of the cold temperature bias, a reduction of the dry bias, particularly over in the Southern Ocean. Model changes possibly responsible for this improvement between these periods are the shift of SSTs from the use of weekly optimally interpolated (OI) SST to the high-resolution real time global (RTG) SSTs and the update of the Community Radiative Transfer Model (CRTM).

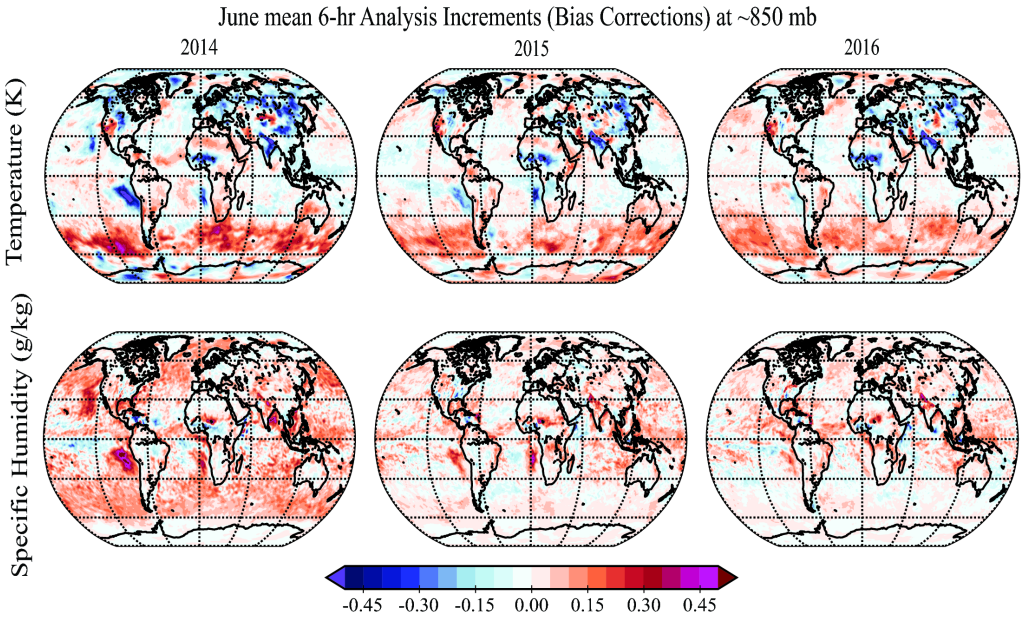

Figure 2.5: Temperature and specific humidity AIs for June 2014, 2015 and 2016. The errors are substantially reduced from 2014 to 2015 especially over the ocean, and further reduce in 2016.

We compared the difference between the high resolution daily RTG and weekly OI SSTs with the changes in AI in 2014 and 2015. In the Northern Hemisphere the surface temperature AI improvements are highly correlated with the places of

significant difference between RTG and OI SSTs (Fig. 2.6). The warmer RTG SSTs in the north Pacific and Atlantic tend to remove the cold bias in 2015, which was found in 2012-2014. Further experiments are required to confirm the role of SST in improving the bias.

In contrast to the situation in the Northern Hemisphere, RTG SSTs are colder in the Southern Ocean. But we still find a reduction of cold bias in forecasted temperature. This is a result of updating the CRTM which improved the analysis of near surface temperature over water, especially in the Southern Oceans by improving specification of microwave sea surface emissivities (http://www.nws.noaa.gov/om/notification/tin14-46gfs_cca.htm,D. Kleist, pers. comm., 2017).

2.6 Periodic Error: Diurnal Cycle bias

Power spectrum analysis revealed that the periodic bias at sub-monthly periods is dominated by the daily cycle, which includes stationary components, a large diurnal component that progresses westward following the motion of the Sun and a significant semi-diurnal signal (Fig. 2.7). The size of these are comparable to the seasonal bias, thus making correction of diurnal and semi-diurnal bias also critical to improving the model performance. To separate these components, we conduct a standard EOF analysis of the 6-hourly AIs each month and then focus on those terms associated with the daily cycle.

Over the eastern Atlantic and Pacific Oceans, the model tends to overestimate humidity and underestimate temperature during daytime and underestimate

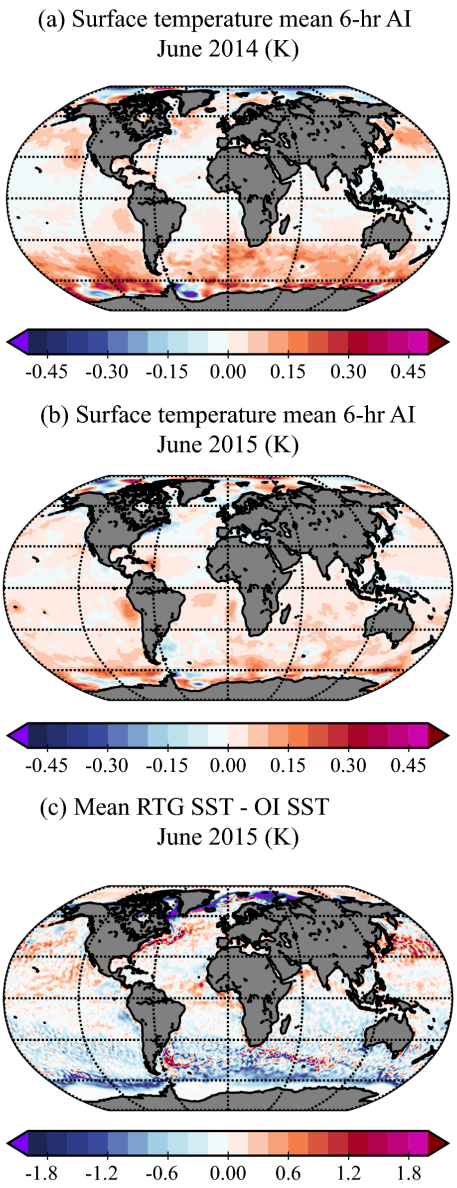

Figure 2.6: Comparison of change in surface air temperature mean bias (K) , (a) June 2014 (b) June 2015 with the (c) difference in RTG and OI SST (K). Warm colors indicate that RTG SSTs are warmer than the OI SSTs.

night-time humidity and overestimate night-time temperature. The bias has a semi-diurnal component during the southwest monsoon season JJA over Europe and Asia,

with peaks in cold bias both in early morning and dusk and warm bias late morning and night.

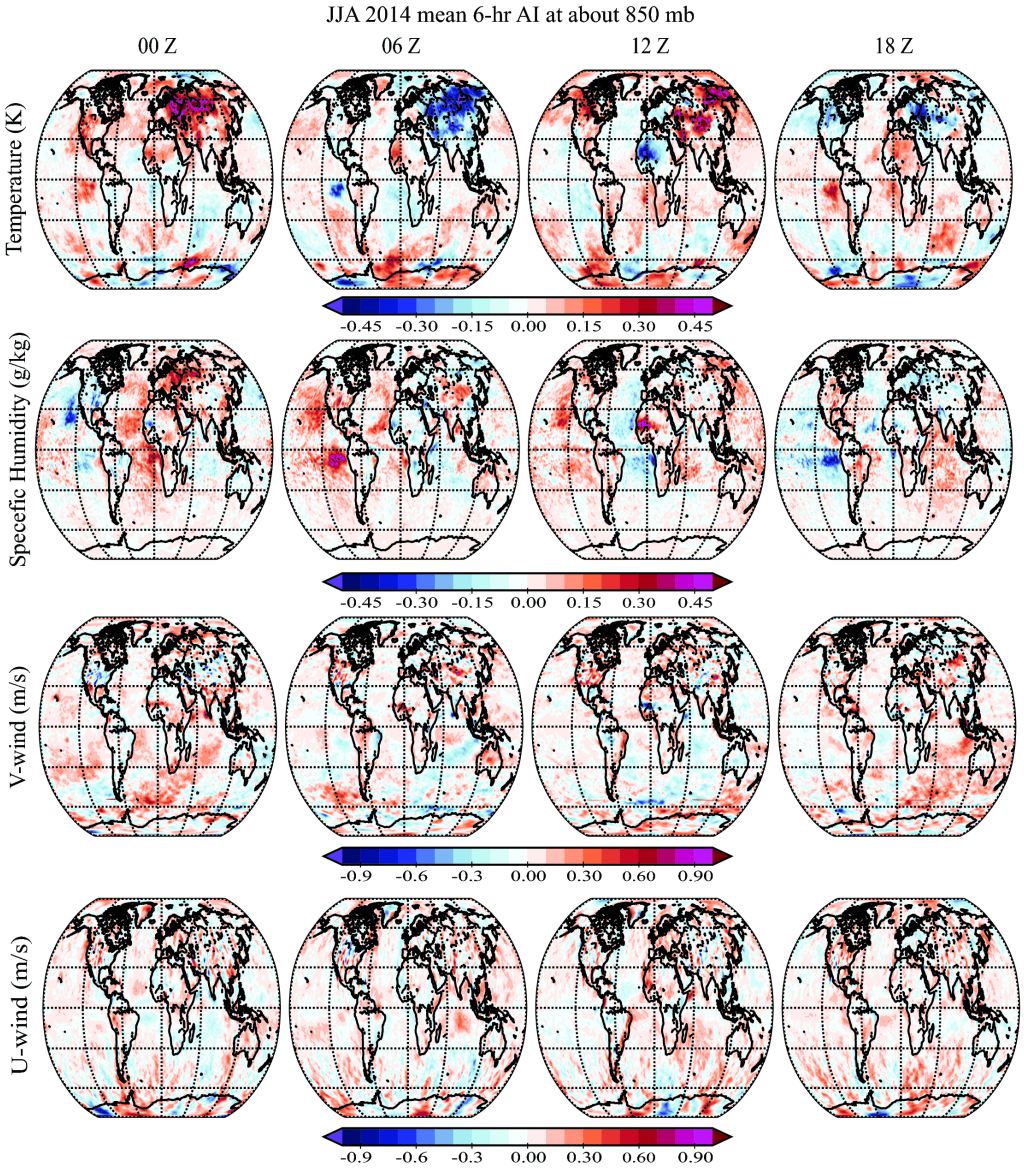

Figure 2.7: JJA AIs for 2014, at 00 Z to 18 Z (from left to right) for temperature, specific humidity, zonal and meridional winds (top to bottom) at approximately 850 mb.

The monthly EOFs, which consist of 120 modes, are dominated by the four leading daily modes which explain 24% (surface pressure), 11% (temperature), and 10% (humidity), and nearly completely describe the daily cycle (Fig. 2.8). The diurnal cycle biases in 2015 and 2016 show similar structure with reduced magnitude.

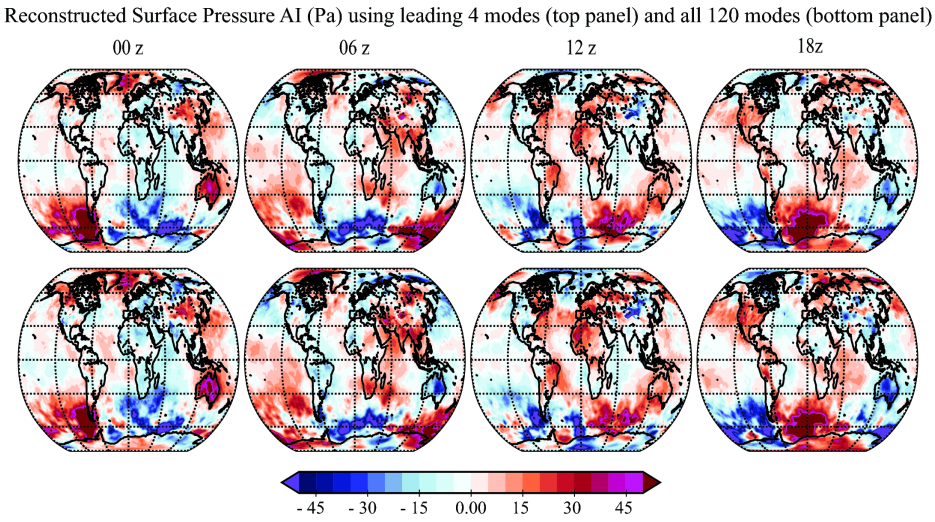

Figure 2.8: Comparison of the diurnal cycle (September 2014) constructed using the first four modes (top row) with the total diurnal cycle (bottom row) errors at 00Z, 06Z, 12Z and 18 Z (left to right) for surface pressure (Pa). This is also true for other variables in different months (not shown).

2.7 Summary and Conclusions

In this chapter, the model deficiencies in the Global Forecasting System (GFS) that lead to systematic errors in the forecast were estimated, as a first step towards correcting them online (i.e., within the model) as in DKM07. To do so, we first

examined the accuracy of using analysis as the verifying truth for estimating model biases. We found that for a small time step and over a sufficiently large sample, the model tendency errors can be estimated as the averaged difference between the analysis and forecasts plus an additional β term. The β term represents the discrepancy between using the truth versus using analysis as truth and arises due to the analysis errors and forecast errors due to inaccurate initial conditions. The averaged analysis minus forecasts provides a good approximation of model systematic errors when this β term is negligible. This is true when the observations are unbiased and highly accurate and the model forecasts and observation have Gaussian statistics with negligible nonlinear higher order terms. This method would however under/over estimate model biases when β term is non-negligible due to:

1. Large observation biases, representative errors, errors in observation operator are present.

2. Analysis errors due to the presence of model bias when the accuracy of observation is comparable or less than the model background.

3. Statistic for observation and background errors are erroneously assumed Gaussian.

4. Significant nonlinear error growth.

5. Analysis error propagated by the TLM becomes significant.

The problem of quantifying the cumulative impact of all these terms is complex and out of scope of this dissertation. Despite these limitations, native analysis

generated using the same dynamical model and being present on the required grid and available at required temporal resolution, seems to be a reasonable choice of verifying truth. Assuming linear error growth in first six hours and highly accurate and unbiased observations, 6-hr AIs provide a good first estimate of the model errors.

We then examined the, 6-hr averaged AIs for the years 2012-2016. AIs are the difference between the gridded analysis and forecast, with the former providing our best gridded estimate of the true state of the atmosphere. They contain information about the physical processes that the model lacks and give the best estimate of the systematic errors arising due to model deficiencies. The 6-hr cycle time is sufficiently short that the errors are still linear. This reduces the likelihood of having errors in one variable at one location inducing errors in another variable at a different location, and thus simplifies the identification of causes of the errors.

Our results reveal the presence of significant bias that is geographically anchored with continental scales in the GFS. The model has excess heating and drying of south and east Asia especially during JJA, which leads to a lower pressure forecasts. A likely cause is weaker moisture-carrying monsoon winds from the Southern Hemisphere, which also affects monsoon convection and circulation. Warm and dry anomalies are also present in the regions where GFS is unable to maintain sufficient stratus clouds, i.e. the zone west of South Africa, and the Americas.

At higher latitudes, the oceans have a cold bias during local summer with northward displacement of the band of intense easterlies over the Southern Ocean. The amplitude of the bias declines in 2015, especially over the ocean. We are able to identify one possible cause of the reduction in the Northern Hemisphere, which

was the switch in 2015 to an improved, higher spatial and temporal resolution in the estimation of SST boundary conditions. However, the bias represented by AIs over oceans in 2012-2014 are not completely due to model deficiencies, but also arise from bias in prescribed SSTs and a problem with observational assimilation. The mean bias is also reduced over the Southern Ocean in 2015. In this region, the change in SST has less impact. Instead, we think the reduction in bias is due to updating of the Community Radiative Transfer Model and improvements in radiance assimilation.

In addition to time mean bias, we find strong daily bias in temperature, surface pressure, specific humidity, and winds. Specific humidity has a strong diurnal bias pattern while the periodic component of temperature bias shows a complex pattern, with both semi-diurnal and diurnal components, where polarity changes every 6-hrs at some places and every 12 hours at other places. The daily biases are similar from 2012 to 2014, and can be represented by the four leading EOFs, computed every month, for surface pressure, temperature, and humidity for all months. The amplitude of the daily biases also declines in 2015, especially over the ocean. Here also, we think the decline in bias is due to the improved SST boundary conditions.

Our results for bias estimation in GFS support the application of the methods used by DKM07 to correct the mean and diurnal systematic errors. As the error growth in the short-term is still linear, we can use the estimated model bias corrections and add them as a forcing term in the model tendency equation. With the best estimate of model biases prior to nonlinear growth, the challenge that now arises is how to utilize the past estimates to correct present models. We address this in the following chapter, we plan to use the successful approach of Greybush et al.

(2012), who used the mean of a limited number of past AIs (e.g., the past 15 days) to correct the model online.

We emphasize that the ultimate goal of this study is not to empirically correct the model bias and improve the forecasts only. This approach can then be used to help guide and optimize the design of subgrid-scale physical parameterizations, more accurate discretizations of the model dynamics, boundary conditions, radiative transfer codes, and other potential model improvements that can then replace the empirical correction scheme. Though this method does not determine the exact source of the error, it provide a starting point to investigate possible causes of the error. The methodology we propose, can be also used to efficiently check potential improvements by testing whether they reduce the mean Analysis Increments as expected from their design.

Chapter 3: Correcting GFS bias with Adaptive Online Correction using Analysis Increments

3.1 Overview

The GFS has significant large scale systematic errors as estimated by the 6-hr AIs in Chapter 2. To reduce the systematic errors, several empirical correction schemes have been developed and tested with simple as well as realistic models (refer section 1.3.2). These schemes include both offline, where correction is applied after the forecast is generated, and online correction schemes, where the correction is applied during model integration. A major disadvantage of the online correction schemes is that they allow forecast errors to grow until the end of the forecast cycle before comparing them with the verifying analyses and making an average correction. This nonlinear interaction also obscures the physical origin of errors.

Several past studies have tested and compared these two types of schemes using simple as well as operational models. These studies found that online correction schemes generally perform better than offline correction schemes (for details please refer Section 1.3.2). The methodologies tested so far with GFS present a good analysis for pedagogical use. However, they cannot be implemented online.

For example, Yang et al. (2008) tested a methodology that requires past 21-month data, during which model changes several times. Xue et al. (2015b)'s methodology required running the model 20 times every 6 hours rendering it computationally too expensive.

In this chapter, we introduce an adaptive online correction scheme following the methodology developed by Danforth and Kalnay (Danforth and Kalnay, 2008a,b; Danforth et al., 2007) to empirically correct this time mean component of the systematic GFS error, i.e. model bias during the model integration. This scheme is based on nudging (Saha, 1992), where a forcing term is added to relax model state towards the estimated truth. We implement this methodology to correct the time mean state-independent systematic errors in the GFS. We analyze the scheme's performance by investigating the changes in the RMS error, bias and error variance/unbiased component of error, and the error growth. We also compare the performance of online scheme with that of a similar offline correction.

We begin by briefly reviewing DKM07 methodology and experiments in Section 3.2. Section 3.3 describes the adaptive online correction scheme that corrects the bias due to model deficiencies. The GFS model and the experimental set-up are described in section 3.4.0.1. Section 3.5 presents the results by comparing the RMSE, bias, unbiased component, ACC and error growth. We finally end this chapter by summarizing our conclusions in section 3.6.

3.2 Review of Danforth and Kalnay methodology

Danforth and Kalnay (Danforth and Kalnay, 2008a,b; Danforth et al., 2007) developed and tested an online correction scheme that estimates and corrects the state independent, time mean and diurnal cycle errors, along with the state dependent errors with a simple SPEEDY model (Molteni, 2003). In this section, we present their methodology and briefly discuss their results.

DKM07's approach involved estimating the model tendency errors and then adding the correction terms in the model tendency equation to generate online corrected forecasts. DKM07 used NCEP Reanalysis (Kalnay et al., 1996) (NNR) as the verifying truth. DKM07 separated the model tendency error in three systematic components, the mean bias, diurnal cycle error and state-dependent error, and random errors as given by eqn (2.1). As the work presented here is mainly concerned with the state independent model errors, we discuss DKM07 methodology only for the bias and diurnal cycle error.

1. **Bias:** DKM07 estimated monthly bias using the the eqn. (2.8) where they used NNR as $\mathbf{x_a}$. As SPEEDY is a less sophisticated model compared to the one used to generate NNR, the term α_m, representing the model bias, is large compared to the β term. Thus, the mean difference between the NNR and 6-hr SPEEDY forecasts divided by 6-hr over a large training period can be assumed to accurately estimate the model tendency bias. They attributed the monthly mean bias correction to the 15^{th} of each month and estimated the bias correction term using daily interpolation. For example, the correction to

be applied for a forecast initialized at February 1, was estimated by averaging monthly mean corrections for January and February over the training period. The bias corrected model was then given by

$$\frac{dx}{dt} = \mathcal{M}^{+}(x(t) = \mathcal{M}(x(t) + \frac{< x_{NNR} - x_b^{6hr} >}{6hr} \tag{3.1}$$

where $< . >$ denotes the mean over the training period,

2. **Periodic errors:** DKM07 found significant periodic errors in the SPEEDY forecasts. These errors were dominated by errors in the diurnal cycle, as expected given that SPEEDY lacks any diurnal forcing. They analyzed the Empirical Orthogonal Functions for the anomalous corrections and found that diurnal cycle could be represented using just two leading EOF modes. They then developed a low dimensional approach to correct for periodic errors. The correction term is given by the second term on right hand-side of eqn. (2.1). It is referred to as low dimensional correction because only the time amplitude which has much lower dimension than full model $\beta_l(t)$ in that equation is estimated online. To estimate this online, DKM07 first estimated the time series of $\beta_l(t)$ by projecting the anomalous corrections obtained from the biased corrected forecast and NNR2 on the leading EOFs during the training period. They then calculated the average amplitude at 0000UTC, 0600UTC, 1200UTC and 1800UTC separately. The time-dependent amplitude can then calculated online by interpolating the 6 hourly amplitudes at the current forecast step.

They found that applying these state independent corrections significantly improved the forecast skill as measured by anomaly correlation coefficients. Using this

online correction scheme, they not only corrected bias, but were also able to reduce the random errors significantly. They also found that their online correction scheme performed better than the tested offline correction applied *a posteriori*. However, they listed a major limitation of this methodology that it requires large amount of training data which would be difficult to obtain with the frequently evolving operational models today.

3.3 Adaptive online correction scheme

In this section, we introduce an adaptive online correction scheme following the DKM07 methodology above to correct the bias component of the systematic errors in GFS. The two-key differences of this approach and DKM07 are:

1. **Choice of verifying truth:** Instead of using reanalysis NNR, we here use the native analysis as the verification. the other option available is to use a non-native analysis like ECMWF analysis. However, using analysis generated from a different dynamical model to initialize the model leads to spin-up errors in short terms. As we plan to use the 6-hr forecast to estimate the model tendency bias, this defeats the very purpose. AIs are easily available from the operational model and thus do not require additional computations. As the analysis is generated using the same dynamical model, it does not lead to large spin-up errors. Further, they are available at the required time and grid and hence avoid the complexities and errors that arise from space-time interpolations. However, using native analysis has it's own drawbacks as discussed in

the paragraph following the next point.

2. **Training period:** Contrary to fixed training period used by DKM07, we adopt a moving training period similar to Greybush et al. (2012). To be able to use this methodology real-time to correct GFS, we introduce an "adaptive online correction scheme" that computes the model tendency correction based on the average of previous N days, instead of centering it around the analysis time. This implicitly also corrects for the state dependent errors.

These changes significantly impact the β term in eqn. (2.9). Firstly, as the analysis and background are generated using the same dynamical model, it would be inaccurate to assume that the β term, that includes analysis errors and background errors due to ill-posed initial conditions, is negligible compared to model bias. In absence of verifying truth, estimating the magnitude of β is out of scope of this dissertation. We recommend running OSSEs as a first step to estimate β. For the present study, we assume that highly accurate and unbiased observations are used to generate the analysis and the high order nonlinear terms are negligible. Thus, the β term can be approximated by eqn.(2.14). As the observations are assumed to be highly accurate, one can assume that β is then small compared to the model tendency bias.

Also, using a finite and short training period introduces additional error in the bias estimates using average AIs. The definition of bias assumes averaging over a very large sample. Hence, one would expect using a small sample set would generate a biased estimate of model tendency bias. Several studies discussed in Chapter 1

have shown that a period as small as 7-12 days is long enough to capture model biases in observation dense areas. Additionally, using most recent biases provides a bias estimate which is state dependent. Thus, using a finite moving training period close to the model state should benefit the model forecast more than a long term bias estimate given sufficient unbiased observations are available.

The corrected model can then be given by, neglecting the small β term,

$$\dot{\mathbf{x}}(t) = \mathcal{M}(\mathbf{x}(t)) + \frac{< \delta \mathbf{x}^{ai}_{6hr} >}{6hr} \tag{3.2}$$

Here $\langle . \rangle$ denotes average over past "N" days as in eqn. (3.3). The limits and denominator terms here show $4 * N$ corresponding to the four cycles each day.

$$< \delta \mathbf{x}^{ai}_{6} >= \frac{\sum_{\tau=1}^{4*N} \delta \mathbf{x}^{ai}_{6hr}(t-\tau)}{4 * N} \tag{3.3}$$

The offline correction at different lead times is applied by adding the analysis increments averaged over the moving training period computed at respective lead time of the forecast (eqn. 3.4).

$$\mathbf{x}^{f\_corr}_{\delta t}(t) = \mathbf{x}^{f}_{\delta t}(t) + \frac{\sum_{\tau=1}^{4*N} \delta \mathbf{x}^{\mathbf{ai}}_{\delta \mathbf{t}}(t-\tau)}{4 * N} \tag{3.4}$$

3.4 The GFS Model and Experimental Set up

3.4.0.1 GFS model v2014

The GFS is a three-dimensional hydrostatic global spectral numerical weather prediction model which provides 16 day forecasts produced by the National Centers for Environmental Prediction (NCEP) with current operational resolution of

T1534 from 0 to 10 days and T574 from 10 to 16 days. Chapter 2 provided a brief summary and reference website links for more details about the GFS/GDAS system and their evolution. In this paper, we test our adaptive online correction (AOC) scheme using the GFS v2014 available on Supercomputer for Satellite Simulations and Data Assimilation Studies (S4) (Boukabara et al., 2016). The 2014 version of the GDAS/GFS supports both the eulerian and semi-Lagrangian physics. The configuration used is the Semi-Lagrangian with physics computations done on a reduced linear Gaussian grid with physics time-step of 10 minutes (some of the physics calculation require a smaller time-step and use a time-step of 5 minutes). The GDAS is based on a hybrid 3DVar-ENKF data assimilation system (Wang et al., 2013), which makes use of a background error estimate from a combination of a lower resolution ensemble Kalman filter and a static background error. In our experiments, we use the eulerian time step and grid resolution of T670. The experiments are conducted using a spatial resolution of T670. The analysis is produced using a lower resolution of T254 for the ensemble runs. The model uses 64 hybrid sigma-pressure levels in the vertical, defined as $p(x, y, t) = \sigma_1 * p_s + \sigma_2$ so that they become parallel pressure levels at high altitudes, σ_1 and σ_2 are given parameters, and p_s is the surface pressure. The lower SST boundary condition over the oceans is provided by the RTG SSTs (Thiébaux et al., 2003).

3.4.1 Experimental Set-Up

To implement the adaptive online correction scheme introduced in section 3.3, we first run a control GFS/GDAS run (exp_cnt) initialized every 6 hours from April 30, 2015 18Z to June 14, 2015 18Z. This produces forecast every 6 hours up to 5 days and analysis every 6 hour. The 6-hr forecasts and analysis are then used to calculate the bias corrections as in eqn. (3.3). These corrections are then used in the online corrected experiments to correct the forecasts from June 1, 2015 to June 7, 2015. Four different online corrections are run with training periods of past 7, 14, 21 and 28 days respectively. The correction experiments run only the GFS cycle which is initialized using the analysis from exp_cnt every 6 hours. The details of the experiments are described in table 3.1.

The results from these online corrected experiments are also compared with a simple statistical offline correction. The offline corrections at different lead times are calculated using eqn. (3.4) and applied to the forecasts obtained from the exp_cnt *a posteriori*.

3.4.2 Choice of training period

As discussed in Section 3.3, on one hand a large training period ensures less sampling error in estimate bias and on the other a training period of sufficient most recent samples ensure more relevant estimate to the state being corrected. To address these issues and find an optimum training period, we begin by computing the correlation between the bias correction (i.e. correction we want to apply given

Table 3.1: Experimental Set-up

Exp. name	Training period	Experiment details
exp_cnt	no correction applied	GFS-GDAS run from May 1 00z, 2015 to June 14 2015
exp_corr07	n = 7	Free online corrected forecast initialized every 6 hours from analysis of exp_cnt run from June 1 to June 7, 2015
exp_corr14	n =14	same as above
exp_corr21	n =21	same as above
exp_corr28	n =28	same as above

by the RHS term for eqn. (3.3)), computed with training periods ranging from past 7 to 21 days, with the average AIs calculated for June 1-June 7, 2015 at different forecast lengths (6, 24, 72 and 120 hours) from exp_cnt run. These correlations help assess how good the estimate of bias correction calculated from past is for correcting bias for the next 7 days.

The bias correction forcing calculated for thermodynamic tested variables, i.e. temperature (left panel of fig. 3.1) are strongly correlated with the future bias for lead times of 1-day and shorter. At these times, the training length is inversely related to the correlation values. Thus, training period as short as 7 days is able to provide a good estimate for bias correction. The correlation values between the wind bias correction forcing and future wind bias (right panel of fig. 3.1) are strong only

for short lead times of 6 hours and decay as lead time increases with very weak to no correlation at 5 days. For longer forecast lengths, the bias correction term for all, q,t,u and v, is only weakly, if at all, correlated with the future bias except near top model levels. This might be due to the rigid top assumption constraining the errors on the top and retarding the nonlinear growth there. Though weak, bias forcing calculated with longer training periods tends to have slightly larger correlation with future bias at lower levels for longer lead times. Thus, $N = 7$ seems to be the best choice, at least for lead times 1-day and shorter. However, estimating bias based on very short training period runs the risk of erroneously accounting for some non-bias component which would not be present in long term mean as bias here. This can result in larger forcing term than required and thus may eventually make the model unstable. To address this issue, we tested 4 different training periods of length 7,14,21 and 28 days (Table 3.1) for the online correction.

3.5 Results

In this section, we analyze the impact applying adaptive online correction (AOC) scheme has on the time mean component of GFS systematic errors, i.e. the bias, and compare that with that of applying a similar statistical offline correction as usually performed operationally. We begin by comparing the AIs for the control and online corrected experiments with different training periods as mentioned in table 3.1 at forecast lengths ranging from 6-hrs to 5-days. We also examine if the reduction of bias affects the apparent random errors, as described by eqn. (1.1) by

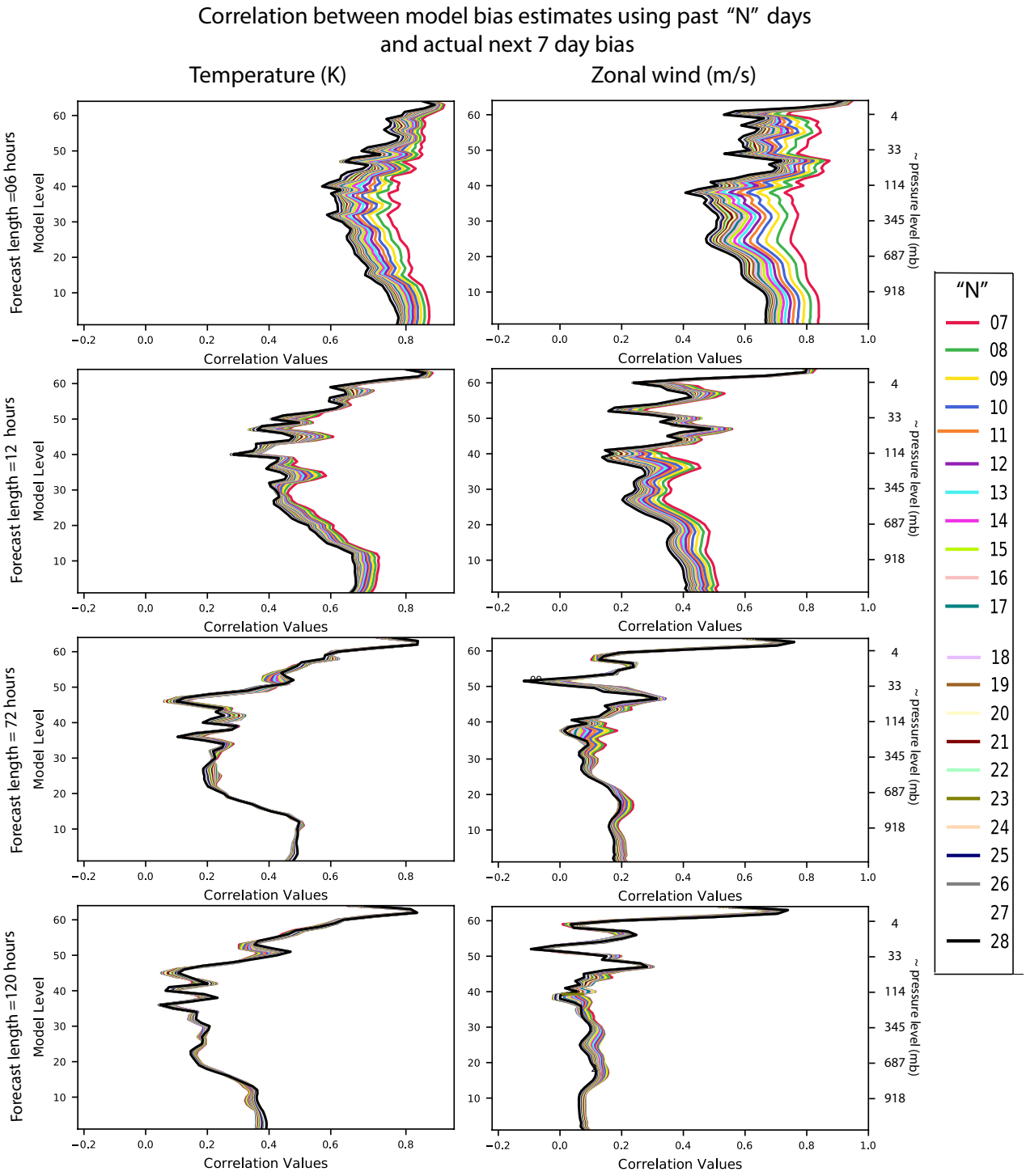

Figure 3.1: Correlation between the different estimates of temperature (left) and zonal wind (right) bias correction from past, averages for training periods ranging from 7 days to 21 days, with the average AIs for June 1-June 7, 2015 from exp_cnt.

inspecting the non-constant term of the model error. And lastly, we compare the improvement achieved by our online correction scheme with the offline corrected forecasts from the exp_cnt experiment, as given by eqn (3.4).

3.5.1 Impact on bias

In this section, we asses the performance of our Adaptive Online Correction (AOC) scheme by comparing the globally averaged, zonally averaged and the global RMS bias in corrected experiments with exp_cnt run. We begin by examining the biases in specific humidity, followed by temperature and the winds.Our experiments showed that the AOC scheme was remarkably stable at all lead times tested. The added correction as forcing term in the tendency equation never led to model blow-up or divergence.

Specific humidity biases occur mainly over the tropics. Biases also start to develop near the pole of summer hemisphere (fig. 3.2) as the model is integrated further in time. The largest density of biases is present above equator and tropics between about 920-700mb (model levels 10-20). The maximum global mean bias occurs at about 850mb and grows from 0.08g/kg at 6-hr to 0.6g/kg at 5-day. This pattern of small initial bias which increases with integration time indicates presence of model deficiencies leading to moist/dry bias. All the experiments with online correction were able to reduce the biases over equator and northern tropics and polar region by about 20% (at 5-day) to 50%(at 6-hr). The ability of the AOC scheme to correct bias in the northern mid-latitudes and most of Southern Hemisphere

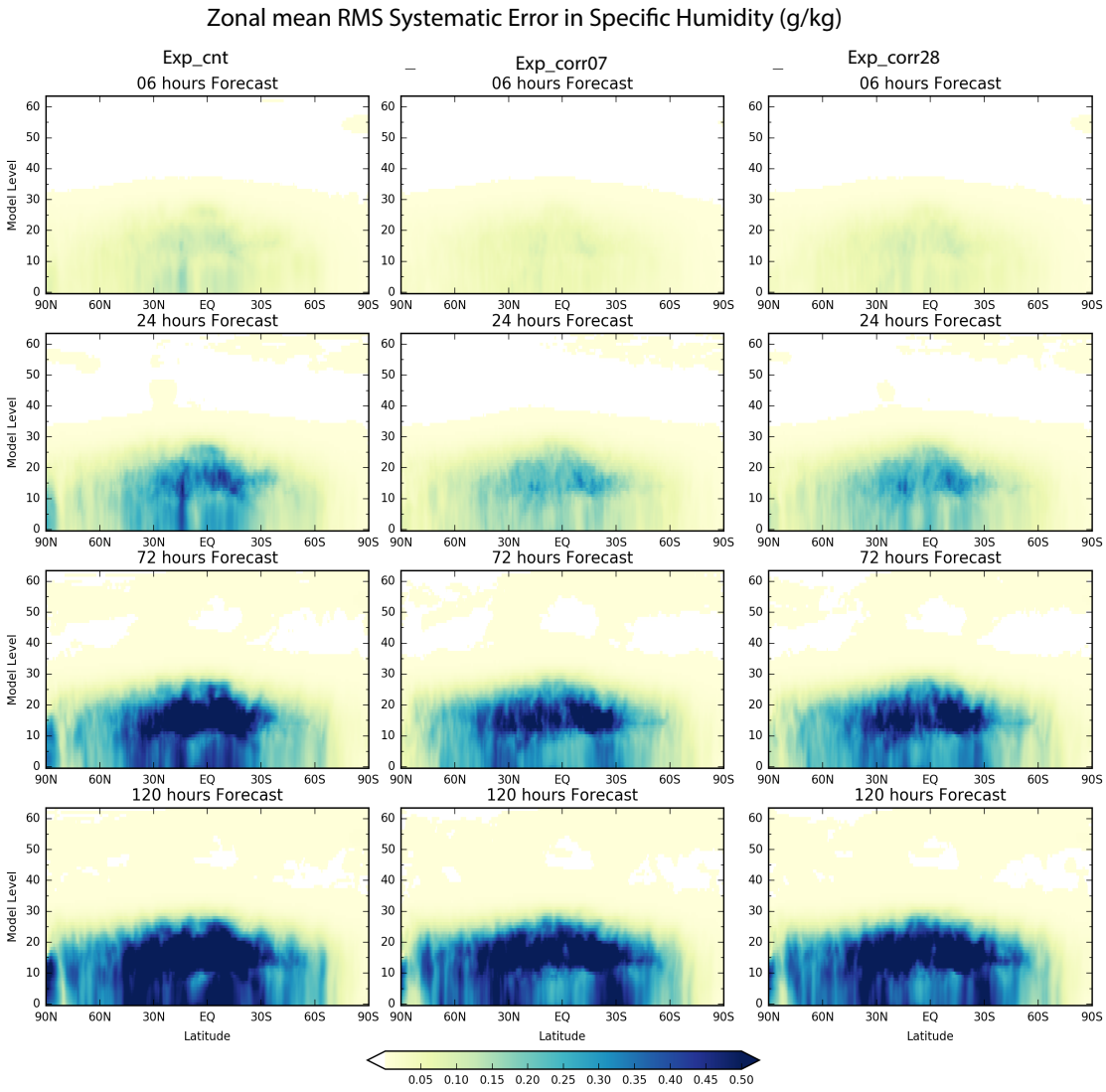

Figure 3.2: Zonal mean RMS bias in Specific Humidity (g/kg) calculated for June 1-June 7, 2015 for exp_cnt(left) and exp_corr07 (center) and exp_corr28 (right).

(south of 30°S) reduces after 1 day (fig. 3.4). Exp_corr07 generated the forecasts with least error. Significant amount of bias in the tropics arises due to model's known inability to maintain stratus clouds. To verify this, we analyzed the zonal mean analysis minus forecast for cloud liquid water at different forecast lengths. As mostly the tropical stratus clouds would contain liquid water or super-cooled water, this can be used as a first step to deduce the lack of stratus cloud. We found that the GFS forecasts underestimated the tropical low-level cloud liquid water content in the tropics at all forecast lengths. This indicates either lack of clouds or discrepancy in the phase in which cloud water is present. The correction scheme is successfully able to identify and correct this bias (not shown here). Dry biases at low levels over South East Asia and oceans north of equator also reduce with all the online correction experiments. The wet bias south of the equator in the Atlantic and East Pacific are also reduced. This might lead to an overall improvement in the ITCZ forecast.

It is also observed that there is generally no additional advantage of using a longer training period of 28 days over shorter training period of 7 days. The specific humidity systematic errors do not show any further reduction as the training period is increased (compare middle and right panels of fig. 3.2). The same is true for all other variables tested (not shown). Hence, moving forward in this dissertation we focus the results from exp_corr07 only.

The online correction has largest impact on the temperature biases. Large temperature biases (about 1K) are present in the mid latitudes and polar region, at low levels (below $\sim$ 900 mb) and near tropopause, and top model levels (fig 3.5).

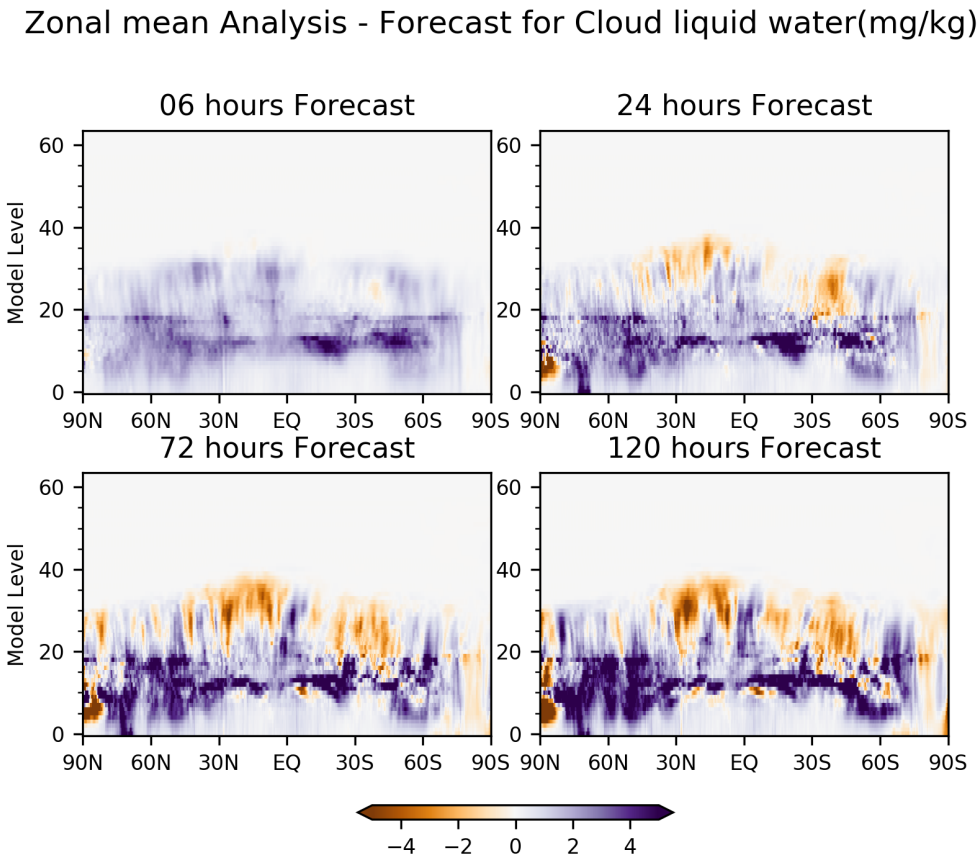

Figure 3.3: Zonal mean Analysis - Forecasts for Cloud Liquid Water at 6-hr to 5 days for exp_cnt.

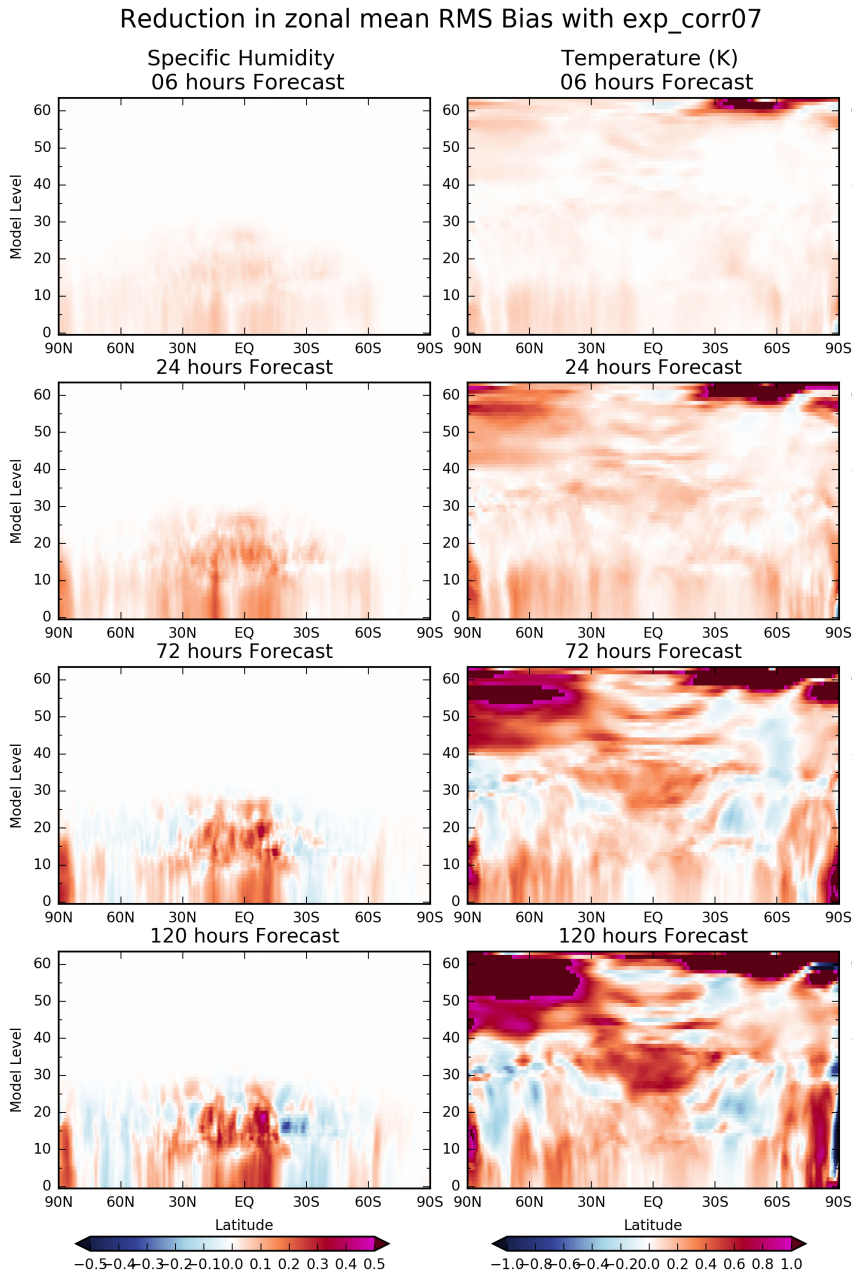

Figure 3.4: Improvement achieved by exp_corr07 in zonal mean bias for at 6-hr to 5-day (top to bottom) for Specific Humidity (left) and Temperature (right). Warmer color indicates that online correction reduced the bias.

Large biases also appears near equatorial tropopause for forecast lengths longer than 1 day. Bias at higher latitudes presents itself as warm bias over Russia and Canada in northern hemisphere and cold bias over Australia, Southern Ocean and parts of Antarctica at lower levels (below 850mb). In the tropics, cold bias is present near surface, which transitions to warm bias above, in the eastern basin of Ocean (fig 3.6), associated with dry bias due to lack of stratus clouds, over Sahara in Africa, and north and central India. Strong biases are also present near model level 30 ($\sim$ 345 mb) in the polar region and model level 40 ($\sim$ 120 mb) in the tropics following the outline of tropopause. This indicates that either GFS generally underestimates the temperature near tropopause or there are errors with estimation of tropopause height.

The online scheme successfully corrects all these biases at all levels for forecast lengths up to 1 day with about 35% reduction in global RMS bias at low levels for exp_corr07 to about 25% for exp_corr28 (fig 3.4). The biases in eastern basins of the oceans, warm bias in Northern Hemisphere and over Africa, and cold bias over Australia are reduced even at longer lead times. This method also completely removes the bias near equatorial tropopause (fig. 3.5 near modle levels 30 (poles) and 40 (tropics)). Even after 5 days, the reduction in global RMS bias is about 20% at low levels (below 850mb), 5% near 500mb and 80% near top model levels. Exp_07 is able to achieve 10% more reduction in biases than other online correction experiments at forecast lengths shorter than 1 day. After that, the improvement achieved by all the online corrected experiments is within 2% of each other, with no one experiment performing better than other at all levels (fig. 3.12).

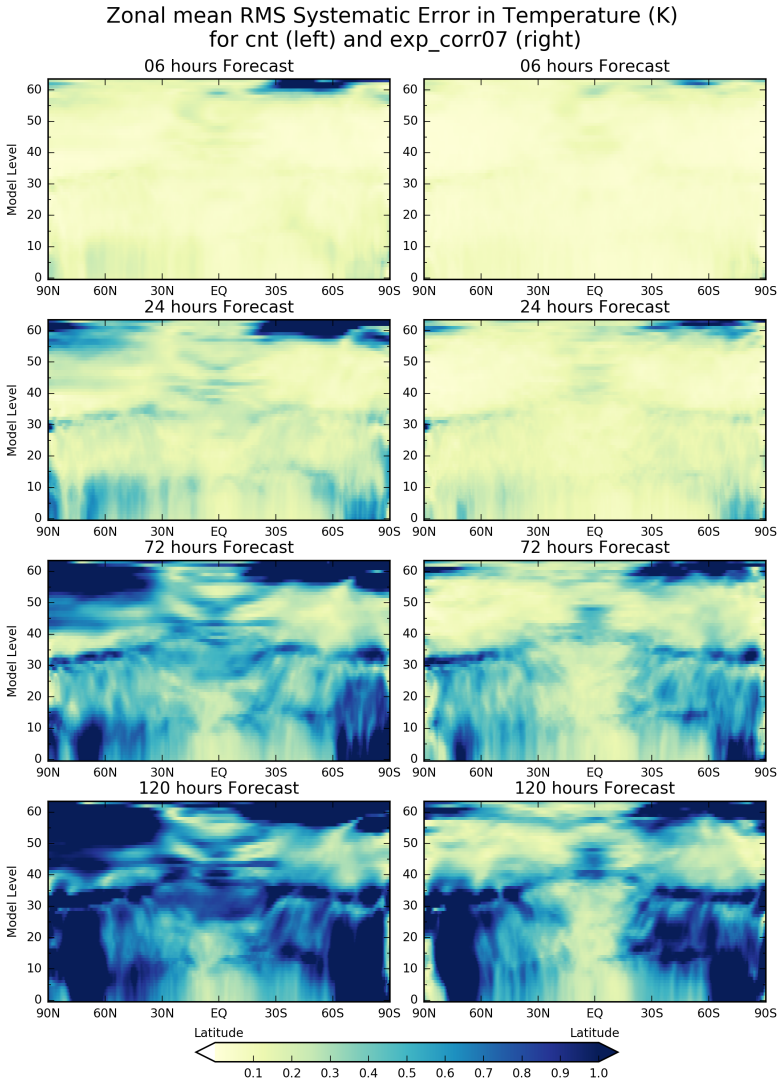

Figure 3.5: Zonal mean RMS bias in Temperature(K) for exp_cnt(left) and exp_corr07 (right)

In some regions, like India and high latitudes in the Southern Hemisphere, despite considerable reduction in biases, small biases still surface after 3 days (fig 3.6). We speculate this is because AOC scheme corrects biases assuming linear error growth even after 1 day. There is a possibility that the nonlinear terms begin to become more significant at larger forecast lengths. Hence, the bias estimated using

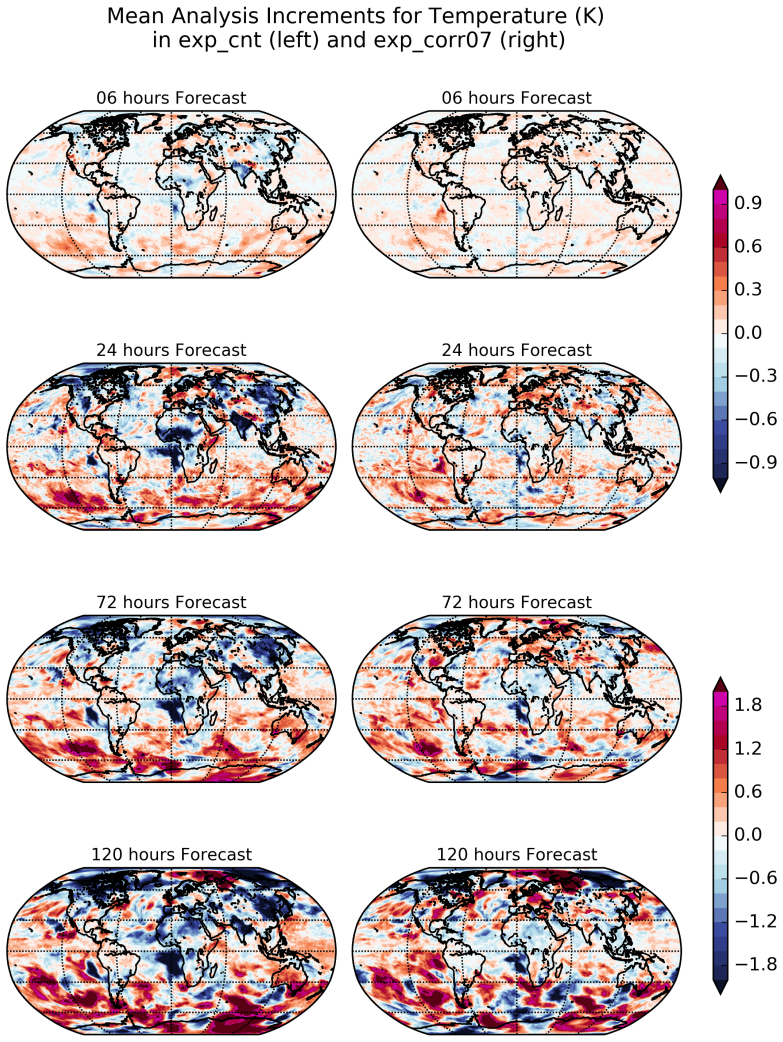

Figure 3.6: Averaged Temperature Analysis Increments calculated for June 1-June 7, 2015 at model level 14 (approx 850mb) for exp_cnt (left) and exp_corr07(right). Warm colors indicate cold bias and viceversa. Exp_corr07 reduces the biases significantly

analysis increments is no longer as good estimate of model bias as it was for shorter length forecasts.

The biases in the zonal winds occur mainly at the tropopause, about 0.3 ms^{-1} (6-hr) to 3 ms^{-1} (5-day), and at top model levels (not shown here). At

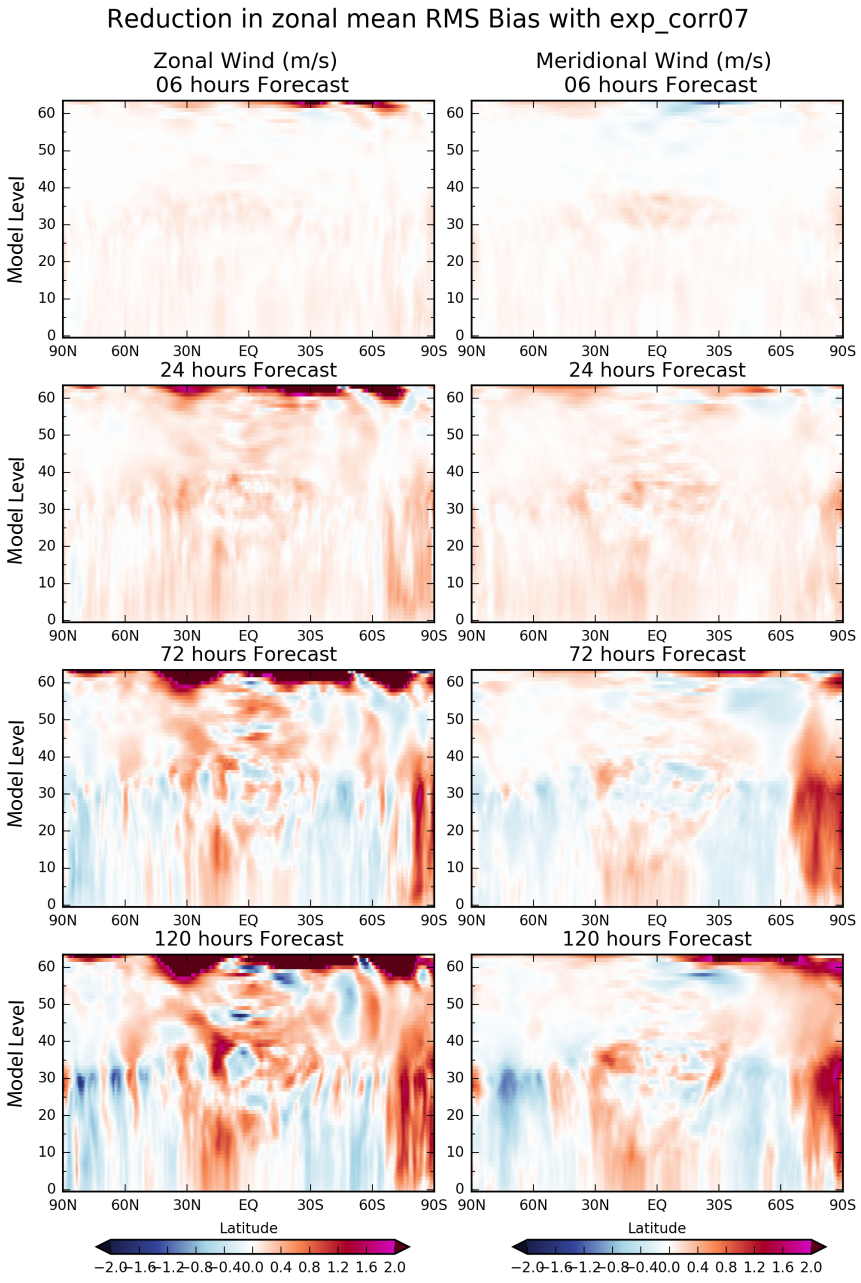

Figure 3.7: Improvement achieved by exp07 in zonal mean bias for at 6-hr to 5-day (top to bottom) for Zonal (left) and Meridional winds (right). Warmer color indicates that online correction reduced the bias.

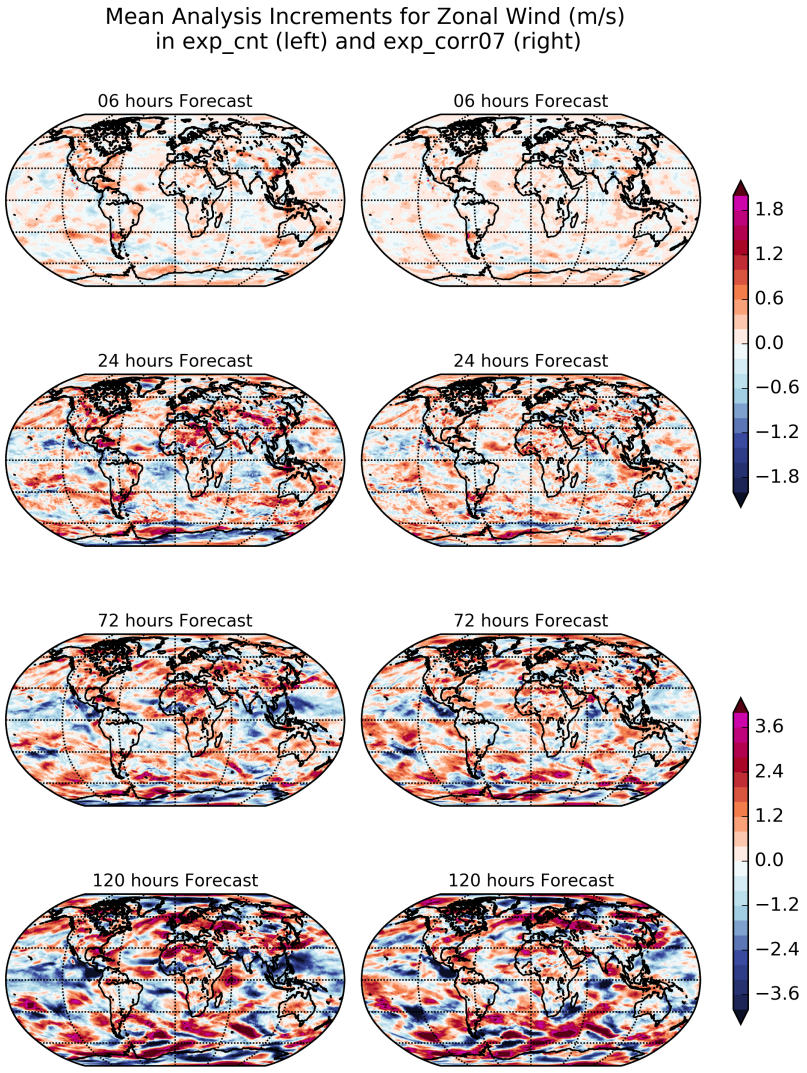

Figure 3.8: Averaged U-wind Analysis Increments calculated for June 1-June 7, 2015 at model level 14 (approx 850mb) for exp_cnt (left) and exp_corr07(right). Warm colors indicate easterly bias and viceversa. Exp_corr07 reduces the biases in the tropics significantly

lower levels, strong easterly bias is present over the islands in Pacific and Atlantic Oceans, mainly, Hawaii, Polynesia, Marina Islands and the Caribbean (fig. 3.8). Westerly bias occurs in the South Pacific off the Argentinian coast (fig. 3.8). After 1 day, westerly bias also appears in the Eastern Pacific near the West coast of the Americas. At the lead time of 5-days, the Pacific is dominated by a strong westerly bias. The online correction experiments successfully resulted bias reduction in the Northern Hemisphere tropics, polar regions of Southern Hemisphere and top model levels by about 40% for lead times as long as 5 days (fig. 3.7). They also lead to reduction in bias by about 45% near 60° north and at level 35 ($\sim$ 250 mb). However, they exacerbate the bias elsewhere when the model is integrated further after 1 day, which leads to a lower percentage reduction in the globally averaged bias. Among the different online correction experiments, the exp_corr07 performs best at lead times of 1 day and shorter, reducing the bias below tropopause by about 40% (global average). For longer forecast lead times, exp_corr28 provides best results at the lower and top model levels, while exp_corr07 still works best at mid-levels. The globally averaged bias below the tropopause generally improves by about 10% after 3-days (exp_corr28) and by only a few percent after that.

The AOC scheme impacts the meridional wind biases in a way similar to that of zonal wind. Most of the biases in the meridional wind forecasts are concentrated near the tropopause, polar regions, and top model levels in the Southern Hemisphere. The magnitude of bias is similar to the zonal winds. At higher latitudes, bias occurs as alternating longitudinal bands of northerly and southerly bias (fig. 3.9). A strong northerly bias is also present near equator near $\sim$ 850 mb over which indicates

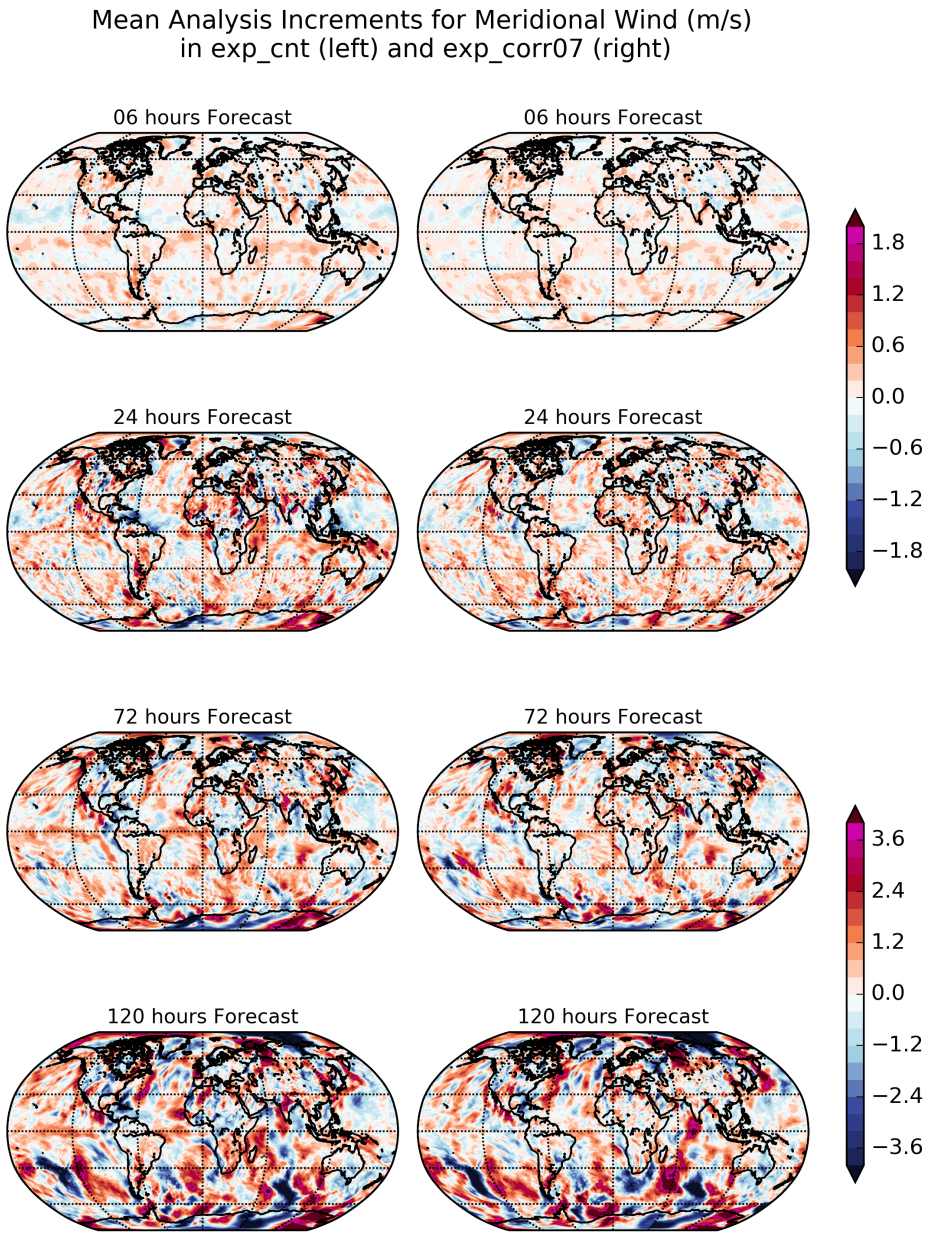

Figure 3.9: Averaged V-wind Analysis Increments calculated for June 1-June 7, 2015 at model level 14 (approx 850mb) for exp_cnt (left) and exp_corr07(right). Warm colors indicate northerly bias and viceversa. Exp_corr07 reduces the biases in the tropics significantly

weakening of the convergence here. At higher model levels, 500 mb, the south tropical Atlantic and Pacific Oceans are dominated by northerly bias. Apart from top model levels, the online method is able to reduce all biases at lead times 1-day and shorter (fig. 3.7). Interestingly, the percentage reduction in 1-day forecast bias ($\sim$ 40%) is more than in 6-hr ($\sim$ 25%). The largest improvements occur at levels below 500mb. For lead times longer than 1-day, the biases over the tropics still reduce by about 40-45% . Though the AOC scheme improves the forecast near South Pole, substantial amount of bias seen in the form of bands still remains. As found previously for other variables, exp_corr07 performs the best at short lead times and after that exp_28 takes the lead.

The least impacted variable by the application of AOC scheme is the surface pressure. The GFS consistently underestimates the surface pressure over North America, north-eastern Africa and South East Asia and parts of Atlantic and Indian Ocean, eastern basin of Pacific Ocean along the coast of Americas, while overestimating it over the oceans at high latitudes generally (fig. 3.10). The online method is able to reduce these biases over South Asia, North America and eastern shores of the Pacific. Though the 6-hr bias over the Atlantic increases with the correction, considerable reduction in achieved in the longer forecast lengths. In other parts, a mixed response to the AOC scheme is found with bias reducing in some regions while increasing at others. However, all the experiments with online correction were able to reduce the global RMS bias at all lead times (Table 3.2). Exp_corr07 generally performed best with about 17%, 34%, 13% and 11% reduction in the global average surface pressure bias at 6-hr, 1-day, 3-day and 5-day respectively.

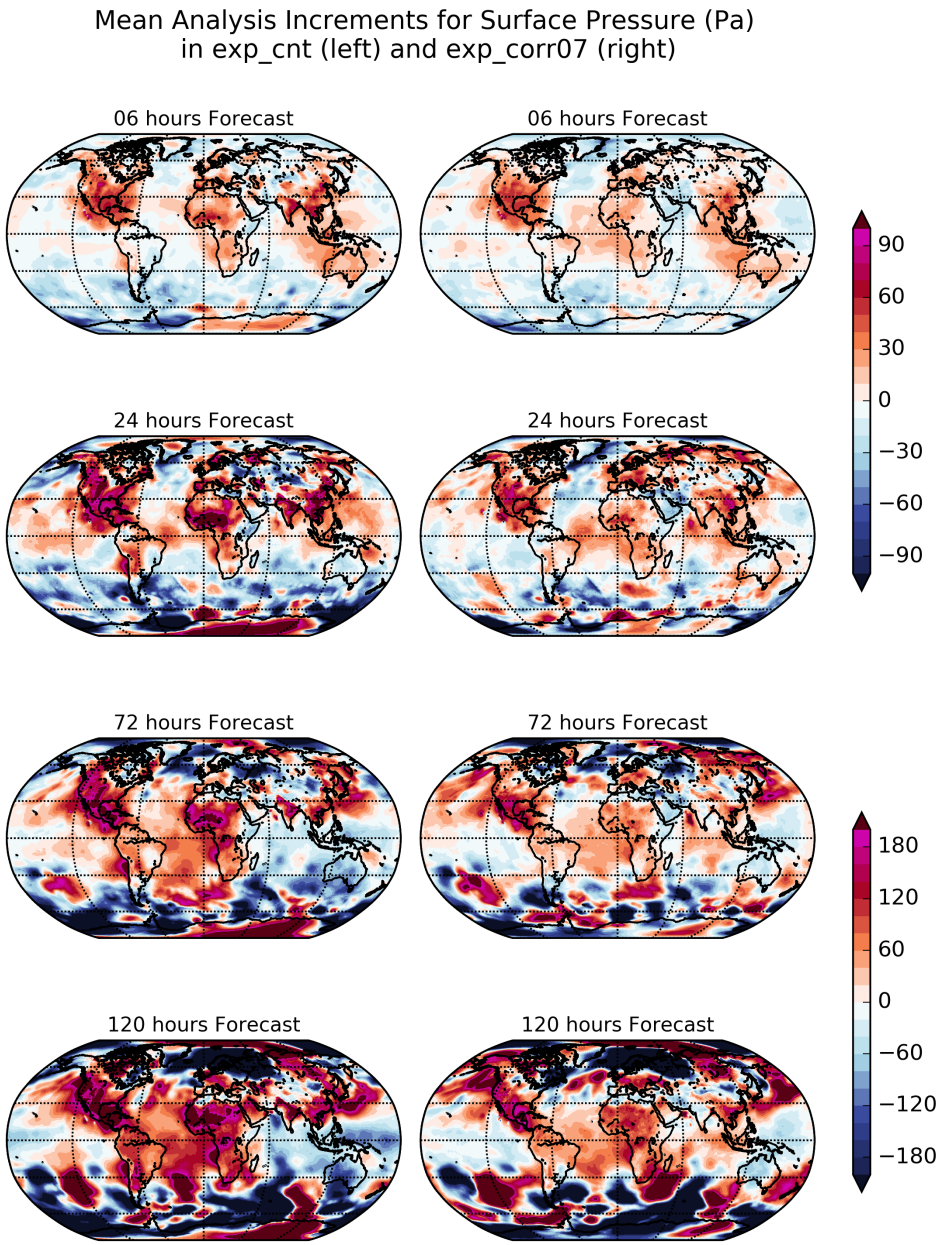

Figure 3.10: Averaged Surface Pressure Analysis Increments calculated for June 1-June 7, 2015 for exp_cnt (left) and exp_corr07(right). Warm colors indicates that GFS underestimates the surface pressure and viceversa. Exp_corr07 reduces the biases over land

We speculate several reasons why the winds and surface pressure are not impacted as well as the thermodynamic variables.

1. Though the corrections to surface pressure are included, no corrections to the cloud/water vapor content are provided. This might lead to an imbalance in the column mass. Thus, this might be a possible explanation that the model surface pressure and winds do not respond as strongly as expected to surface pressure corrections.

2. The corrections are applied to the zonal and meridional winds instead of the vorticity and divergence. It is possible that these wind corrections are not accurately translated to divergence and vorticity corrections in the model. This can lead to a discrepancy between the surface pressure corrections and the corresponding divergence and vorticity changes required to maintain the mass balance in the vertical column.

3. The wind corrections are applied in the GFS physics which takes place on a Gaussian Grid instead of the dynamics in the spectral space. This might lead to information loss as the variables are converted from grid to spectral and vice-versa.

4. The surface pressure and wind corrections are imbalanced. To investigate this, we compared the geostrophic wind estimated from the mean AIs of surface pressure with the mean 10m winds AIs (fig. 3.11). As expected the actual wind increments are smaller than the geostrophic wind increments due to the frictional forces near surface. Even with the reduced magnitude, the wind

Table 3.2: Globally Averaged Surface Pressure Bias averaged over June 1-7, 2015

Exp_name	Lead time = 6 hrs	Lead time = 1 day	Lead time = 3 days	Lead time = 5 days
exp_cnt	15.757	33.159	74.067	121.44
exp_corr07	13.071	21.880	62.428	108.06
exp_corr14	13.662	23.539	60.191	110.36
exp_corr21	13.760	24.068	61.100	111.16
exp_corr28	13.685	24.732	61.341	111.33

pattern for the geostrophic and actual wind components seem to be similar specially near the poles (for example over Greenland and Antarctica). However, the geostrophic and actual U-wind components are not consistent in the South Pacific Ocean. Thus, it is possible that AIs are not balanced everywhere and might impact the AOC scheme. However, more analysis, that is out of scope of this work, is required for a concrete conclusion.

3.5.2 Impact on the Apparent Random Errors

Next, we examine whether the AOC scheme which reduces the bias, i.e. the constant part of the systematic error, also impacts the unbiased component of errors (second term on RHS of eqn.(1.1)). This term includes the periodic, state-dependent and random errors. We found that for lead times up to 1 day, there is no significant impact on the unbiased component of errors, if any at all for all variables tested.

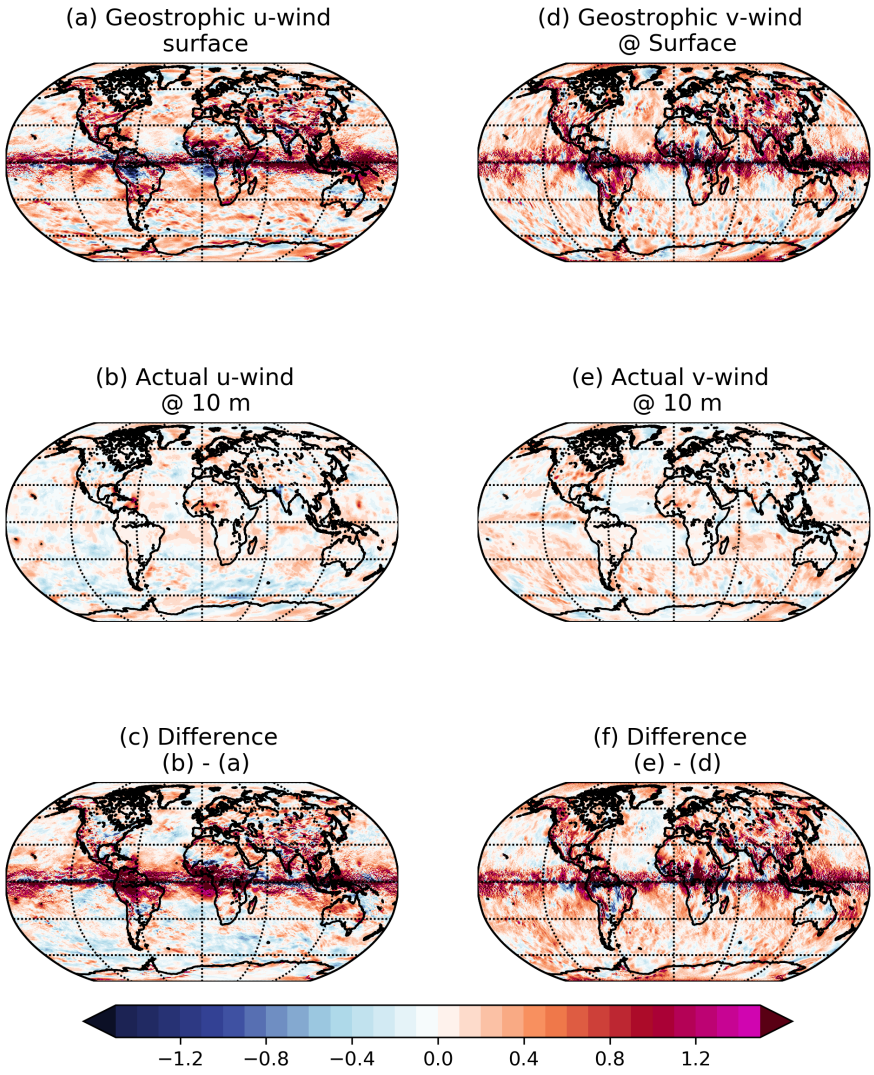

Figure 3.11: Geostrophic u-wind(a) and v-wind (d) estimated from mean surface pressure AIs from June 1-June 7, 2015 compared with the actual 10m u-wind(b) and v-wind(e) and their difference (c) and (f). All units are in m/s.

At longer lead times, the global RMS error generally increase slightly (fig. 3.12). The increase is maximum for exp_corr07 among the different experiments and for temperature among the different variables tested. However, generally this increase is considerably small as compared with the reduction in the bias. Hence, we end up with reduction in the total model error (fig. 3.12). The only exception to this is the error in 5-day temperature forecast near model level 25 ($\sim$ 500 mb), where the improvement achieved in the bias is offset by the worsening of random errors which results in slightly larger total error. This is due to an increase in non-constant errors near the poles at these levels. In the tropics and mid-latitudes, the areas with improvement in bias are closely correlated with improvement or no impact on these errors. Unbiased errors also improve at top levels specially in the polar region. The impact on unbiased errors is very chaotic at low and mid-levels of higher latitudes.

Power spectrum analysis showed that the periodic errors at sub-monthly scales are dominated by diurnal and semi-diurnal errors. Although we are averaging the bias correction over whole days, this mean bias correction can still impact the diurnal cycle errors. When the diurnal cycle errors are in phase with the bias the forecast should improve. However, if they are out of phase, bias correction could worsen the diurnal cycle error making the unbiased component of total error larger than the control run. Hence, this could be one reason that the unbiased errors increase after the correction is applied. This is consistent with the findings of Yang et al. (2008), who also reported increase in GFS random errors when correcting for climatological mean, and Xue et al. (2015a) who inferred that correcting mean bias without correcting diurnal errors lead to worsening of GFS forecast errors.

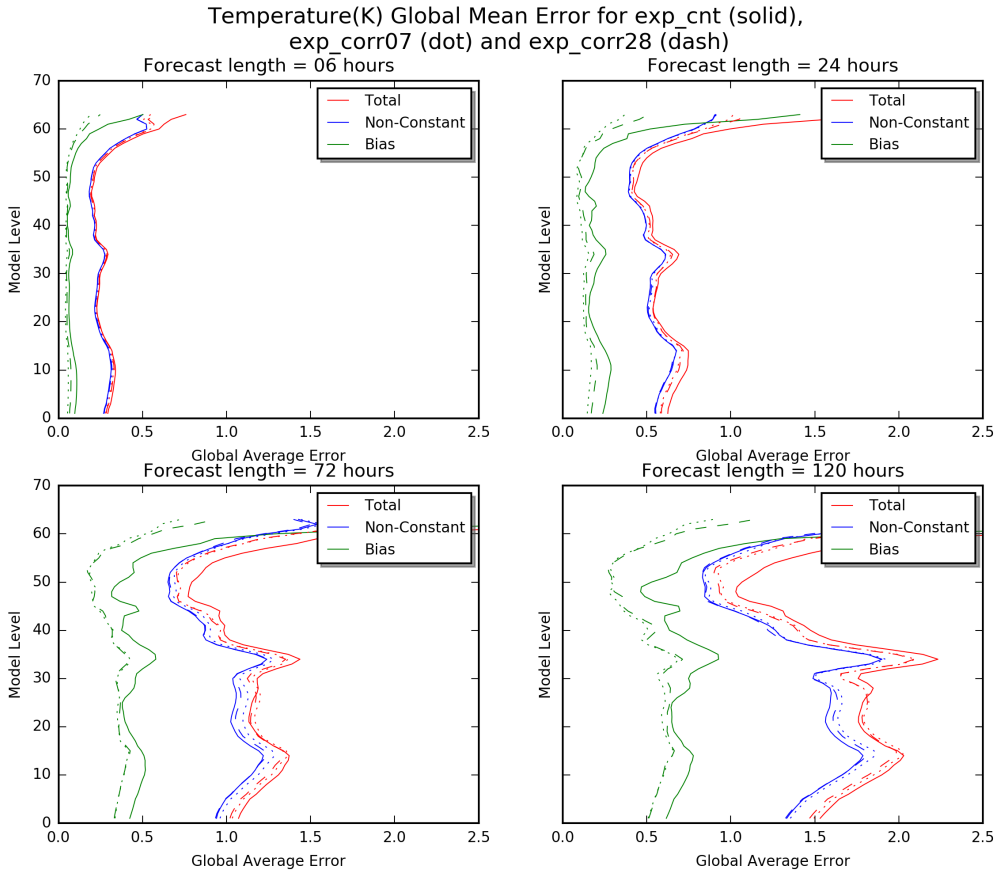

Figure 3.12: Globally Averaged Temperature bias (green), unbiased errors (blue) and total model error (red) averaged for June 1-June 7,2015 for exp_cnt (solid), exp_corr07 (dot) and exp_corr28(dash) for forecast lengths 6 -hr, 1, 3 and 5 days.

3.5.3 Impact on Error Growth

One of the advantages of online correction schemes over the offline schemes is considered to be their ability to reduce error growth. As the correction is applied at each time step, each intermediate state is nudged towards the verification state reducing the error growth. To validate this, we examine the evolution of global mean RMS bias and total error over the full forecast period of 5 days (fig 3.13). In

general, the error growth is reduced in the first 12 hours of model integration. After that the bias grows at a rate similar to that of uncorrected model. Although the error growth is similar, the bias itself is lower because of the slower initial growth when online correction is applied. This is true for almost all model levels except near 500mb. Near this model level the impact on initial bias growth of the wind biases is very small. The biases are smaller for online corrected forecast until about 2 day. After this, the growth rate as well as the bias begin to become higher than the control run.

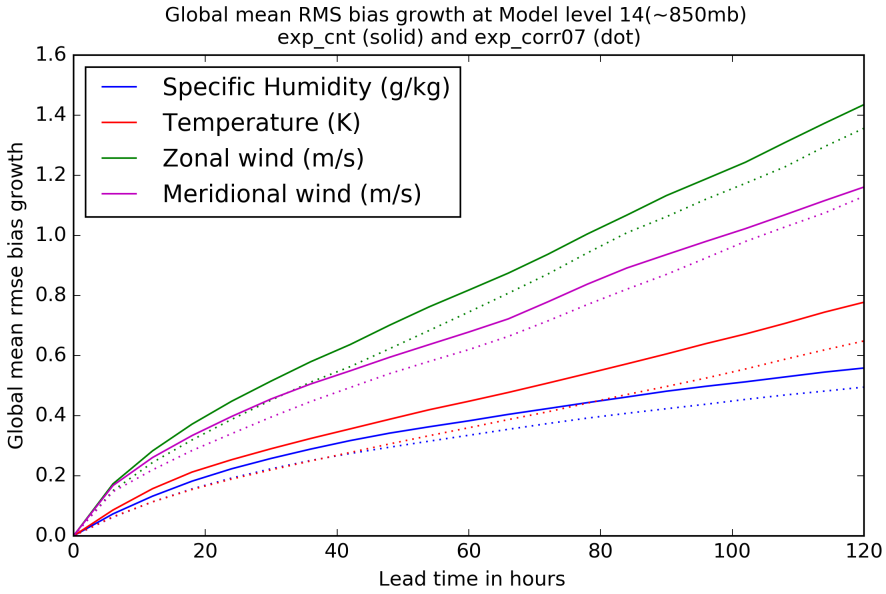

Figure 3.13: Evolution of global mean RMS bias at model level 14 for exp_cnt (solid) and exp_corr07 (dotted)with forecast lead time calculated for the period June1-7,2015. The colors indicate different variables: Temperature(red), specific humidity (blue), zonal wind(green) and meridional wind(magenta).

As the major impact of online correction of error growth is seen in the initial few hours, we ran additional experiments (training period of 7 and 28 days) with a

weighting factor multiplied to the correction term in the model tendency equation. The value of this factor is set to 1 until 6 hours, after which the factor decays linearly such that its value is 0 at 120 hour lead time. This is somewhat similar to the Linearly decaying weighted (LDW) correction applied by (Xue et al., 2015a). They found that using a weighing factor that goes from 1 at the beginning to 0 at the end of forecast cycle, resulted in better short-term GFS forecasts with reduction in wind biases. Our findings, however, differ from those of Xue et al. (2015a). Though the biases in 3-day wind forecasts reduce after the weighting factor is included, we observe little to no change elsewhere (3.14). In fact the large improvements seen at the top model levels for temperature diminish after the application of the weighting factor. Additionally, this also slightly worsened the unbiased component of total error for all variables. Hence, there were not any significant advantages of reducing the magnitude of tendency correction with the increase in forecast length.

3.5.4 Offline vs Online

Finally, we compare the performance of our AOC scheme with the offline correction applied *a posteriori* with the same set of training periods as used for the online corrections. Similar to the online correction, the offline corrections computed with a shorter training period generally lead to the largest bias reduction for short lead times of up to 1-day. Consistent with online corrections the largest bias reduction occurs over the tropics for specific humidity and temperature. The winds are the least improved of all the variables tested with offline scheme as they were for

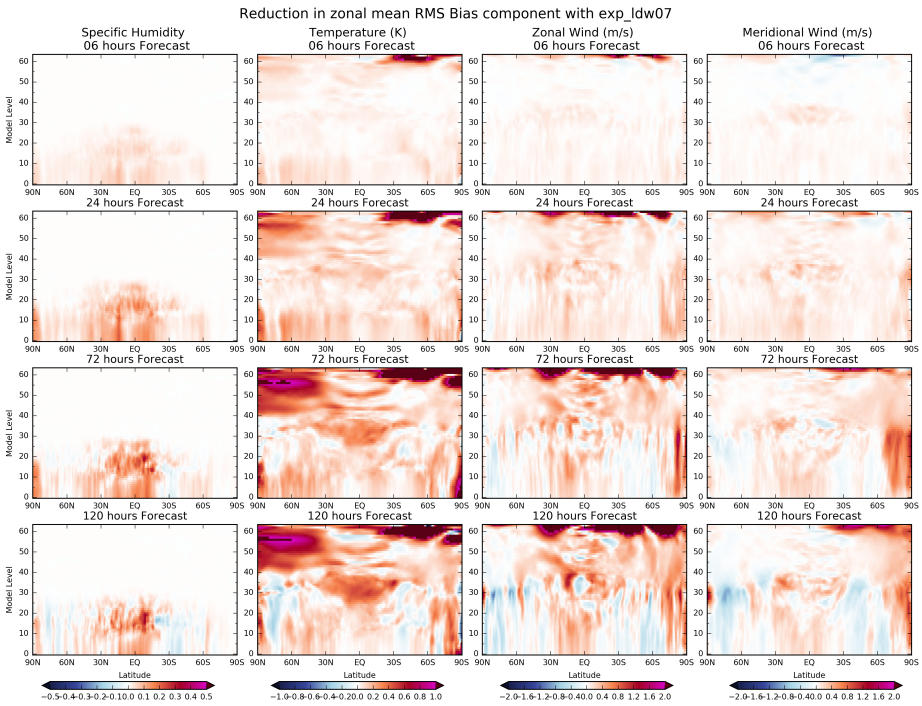

Figure 3.14: Same as fig. 3.4 and 3.7 except a linearly decaying weighting factor is used along with the correction in tendency equation.

the online scheme.

The offline correction applied to the forecast of exp_cnt is able to reduce bias globally for short lead times for all variables and is slightly better than the online correction (fig. 3.15). However, the unbiased component of error increases more than that with online correction. Thus, the improvement in total error is comparable with that achieved by the online correction. At lead times larger than 1 day, the offline correction is not as effective in reducing either the bias or the unbiased component as the online correction. This is especially true for the winds, where the bias increases significantly over the mid and high latitudes after the offline correction is made

(compare fig. 3.15 and 3.7). We found that the global mean RMS bias as well as total error are higher than the exp_cnt for lead times larger than 1 day.

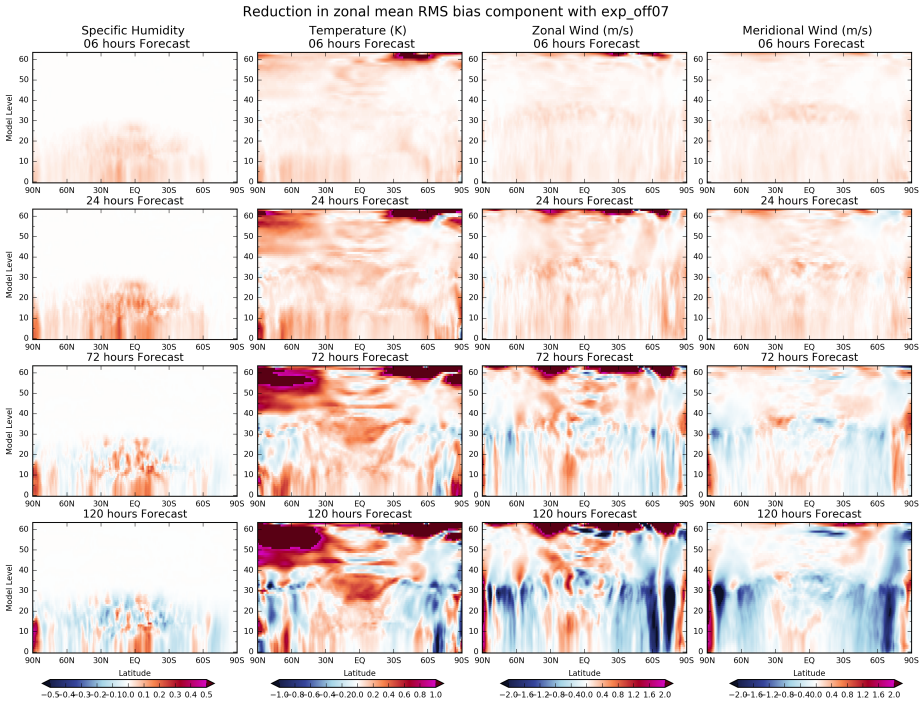

Figure 3.15: Same as fig. 3.4 except offline correction scheme is used instead of online correction.

3.6 Summary and Conclusions

In this chapter, we implemented an adaptive online correction scheme to reduce the bias component of systematic errors in the GFS short and medium range forecasts of surface pressure, temperature, specific humidity, and zonal and meridional winds. This scheme consists of a two step process, (1) estimation of bias corrections based on recent past estimates of the averaged AIs and (2) correcting

for bias by including the correction term in the model tendency equation.

We computed the correction term by averaging the 6-hr AIs over an adaptive training period, which varied from past 7 to 28 days. Though the scheme is based on DK0708, our scheme differs from DK0708 in 2 ways: (1) We here use GDAS analysis as estimate of truth instead of a reanalysis and (2) in place of a training period based fixed independent sample, we here use a running mean for training. A major advantage of using analysis for verification is that, unlike observations, the analysis is available at the required spatio-temporal resolution. Also, AIs are already computed within the GFS/GDAS framework, so that the only additional computational resources required are for computing the average of the AIs. Using a short period of 6 hours, ensures that we compute bias before errors convolute and begin to grow nonlinearly. Hence, 6-hr AIs provide the best estimate of the forecast error before they start to grow nonlinearly. Empirical correction schemes that use fixed training periods of past few years or months require fresh computation of biases with operational model upgrade. Using adaptive training period of past few days overcomes this problem, making it more robust for operational applications.

After estimating the correction term, we correct the model by adding it as a forcing term in the model "physics" tendency equation. Correcting online, instead of an *a posteriori* correction, facilitates the usage of bias calculated at just one lead time to correct forecasts at all lead times. After testing this scheme for different training periods and running forecasts for as long as 5 days, we found running GFS with AOC scheme is remarkably stable. The improvement achieved in the 6-hr forecast was as strong as the average correction applied. This validates our

assumption of linear error-growth during the initial 6 hours. Unlike empirical offline correction schemes which require long training periods, AOC was able to reduce bias by using only past 7 days biases. In fact, the AOC scheme with the shortest training period lead to better forecasts than with longer training periods tested through lead time 1-day. For longer lead times, the choice of training period had only a small impact.

The AOC scheme using only past 7 days for training significantly reduced the biases globally through lead time of 1-day, when the errors are still growing linearly. The globally averaged biases in thermodynamic variables were reduced by about 50-70% and in winds by about 30-45% near and below tropopause and at the top levels. Forecasts in the regions of known model deficiencies, including biases of humidity due to unresolved stratus clouds, temperature biases over South East Asia and wind biases near equator, showed considerable improvement after online correction was applied. The AOC scheme continued to improve the forecasts over tropics below about 500mb and near top levels after 1-day. However, the performance in other areas reduced, even worsening the errors at over mid-latitudes as in the case of wind. One can expect this behaviour as the correction scheme implemented uses a zeroth order correction term and after 1-day we would expect errors due to different sources convolute and start growing nonlinearly.

The AOC scheme, in general, reduced the bias growth in the initial 12 hours, except near 500 mb for the winds. The bias growth afterwards is similar to that of the uncorrected model but the bias itself is significantly smaller. As the error growth slows down after initial few hour, we tested correcting online using linearly

decaying weights. Contrary to the results in past studies Xue et al. (2015a), we found that there wasn't any advantage of using this decaying weighting factor. The only significantly impacted forecasts was of 3-day winds for which the the magnitude of bias reduction in the areas improving increased and the worsening of bias decreased over the mid latitudes. No significant improvement was observed in the growth to total error.

We also found that the online correction performed significantly better than the offline correction at lead times larger than 1 day. Though the global mean RMS bias reduction achieved by the offline correction was larger than the online correction at lead times shorter than 1 day, the random error were much worse. Thus, the impact on the total error at these lead times was similar for both online as well as offline scheme.

Chapter 4: Summary and Future Directions

4.1 Summary

In this dissertation, we successfully introduced and tested an adaptive online correction (AOC) scheme that can be implemented in real-time to estimate and correct the biases present in the NCEP's operational GFS model (v2014) during the model integration. This scheme consisted of a two-step process, (1) estimation of bias corrections based on recent past estimates and (2) correcting the bias online by adding correction terms of the tendency equation.

In chapter 1, we first developed a theoretical framework to assess how model systematic errors can be estimated using analysis increments and what are the limitations of such technique (summarized in section 4.2). We then estimated the state-independent component of GFS systematic errors using 6-hr Analysis Increments (AI) as a first-step towards correcting the GFS online. The ultimate goal here is to estimate the model tendency corrections using these error estimates, that can then be added to the model equation to obtain the online corrected model. Assuming that the error growth is linear during the first six hours and the observations are bias corrected before being assimilated to generate analysis, 6-hr AIs provide the best estimate of systematic errors arising due to model deficiencies which can later be

used to estimate tendency correction. Estimating model error as the negative of 6-hr AIs, we found that significant large scale biases that are geographically anchored are present in the GFS. These biases stay consistent for the period of 2012-2014. In addition to the time mean bias, we found strong diurnal and semi-diurnal biases in temperature, surface pressure, specific humidity, and winds. Specific humidity has a strong diurnal bias pattern while the periodic component of the temperature biases show a complex pattern, with both semi-diurnal and diurnal components, where polarity changes every 6-hrs at some places and every 12 hours at other places. The daily biases are similar from 2012 to 2014, and can be represented by the four leading EOFs, computed every month, for surface pressure, temperature, and humidity for all months. The amplitude of the time mean and daily biases declined in 2015, especially over the ocean. This was attributed mainly to the improvement of SST boundary conditions and the data assimilation system.

In chapter 2, we introduced an adaptive online correction scheme following the methodology of DKM07 that uses 6-hr AIs to correct the GFS model online. Contrary to DKM07, this AOC scheme use GDAS analysis as estimate of truth instead of a reanalysis and in place of a training period of fixed independent samples, and estimates tendency correction by averaging the 6-hr AIs over an adaptive training period, which varied from past 7 to 28 days and dividing it by 6-hr. After estimating the correction term, we correct the model by adding it as a forcing term in the model tendency equation. Correcting online, instead of *a posteriori* correction, facilitates the usage of bias calculated at just one lead time to correct forecasts at all lead times.

The online correction reduced the globally averaged biases in the short term forecasts of thermodynamic variables by about 50-70% and in winds by about 30-45% near and below tropopause and at the top levels. Forecasts in the regions of known model deficiencies, including biases of humidity due to unresolved stratus clouds, temperature biases over South East Asia and wind biases near the equator, showed considerable improvement after online correction was applied. The AOC scheme continued to improve the forecasts over the tropics below about 500mb and near the top levels after 1-day. However, the performance in other areas reduced, even worsening the errors over the mid-latitudes as in the case of wind. The results also indicated that the online correction scheme is able to reduce the bias growth during the initial linear error growth phase. After a few hours, bias grew at the same rate as the bias in the uncorrected model but had a lower saturation value. The total error growth showed negligible reduction with the online correction. We also offered possible explanations for little impact of AOC on winds and the surface pressure.

We now revisit the scientific questions raised in Chapter 1.

4.1.1 Can a stable online correction scheme that is appropriate for operational use be designed?

As found in previous research the online schemes have a tendency to blow up the model if not designed properly. The AOC scheme implemented here remained stable during the model integration time tested, i.e. 5-days and reduced the

systematic errors significantly.

The scheme uses 6-hr AIs which are already available operationally at no additional cost. Unlike several other online correction schemes that require long training periods (like DelSole et al. (2009, 2008); Yang et al. (2008)) or several model runs (as in Xue et al. (2015a,b)), the AOC scheme was able to reduce bias by using training period of just 7 days. Though using a larger training period of 28 days led to a larger reduction in 3 to 5-day forecast, the advantage was only marginal. Hence, the computational costs associated with estimating the tendency correction terms in the operational setting are very low.

4.1.2 Can short term model error, based on analysis increment, be used to represent model tendency errors?

Estimating tendency corrections (negative of tendency error) requires calculating forecast errors during a short integration period, where the errors are still growing linearly. In the absence of true atmospheric state, we used 6-hr analysis increments which are the gridded differences between GDAS analysis and the 6-hr forecast to estimate the systematic correction (negative of systematic error). Assuming that the error growth is linear during the first six hours, we estimated the bias correction terms by dividing the average 6-hr AI by 6 hours. The improvement achieved in the 6-hr forecast was as strong as the average correction applied using 6-hr AIs. This validates our 6-hr linear error growth assumption. Thus, large scale model tendency errors can be estimated using 6-hr forecast and a verifying truth.

However, it should be acknowledged that averaged analysis increments, apart from the model biases, also contain the mean analysis errors and model errors due to inaccurate initial conditions (discussed in section 4.2. Thus, it's likely that they lead to under/overestimation of analysis increments.

4.1.3 What are the general characteristics of model error and error growth in GFS?

The GFS has significant large scale errors that remain geographically anchored through 2012-2014. Specific humidity biases occur mainly over the tropics. Biases in specific humidity develop at high latitudes of the summer hemisphere as the model is integrated with time. Large temperature biases (about 1K) are present in the mid latitudes and polar regions at low levels (below900 mb) and near the tropopause, and top model levels. The biases in the winds occur mainly at the tropopause, about 0.3 ms1(6-hr) to 3 ms1 (5-day), and at top model levels. In general, the errors grow steeply during first 6-12 hours and the growth slows down after that. The error growth in the first 6 hours is almost twice as much of the error growth from start to 1 day. The model has excess heating and drying over south and east Asia especially during JJA, which leads to an erroneously lower pressure forecasts. A likely cause is weaker moisture-carrying monsoon winds from the Southern Hemisphere, which also affects monsoon convection and circulation. Warm and dry anomalies are also present in the regions where the GFS is unable to maintain sufficient stratus clouds, i.e. the zone west of South Africa, and the Americas. At higher latitudes, the oceans

have a cold bias during local summer with northward displacement of the band of intense easterlies over the Southern Ocean.

Apart from time mean errors, the GFS has significant periodic errors that are dominated by the diurnal cycle.

4.1.4 Can an online correction aimed at correcting systematic error also impact the random component of model error?

Applying online corrections led to a slight increase in the unbiased component of the error, referred to as random errors in eqn. (1.1). However, it can be argued that this increase is due to the absence of a diurnal cycle correction to the GFS and not a true representation of an increase in random errors. Corrections of the mean bias at times where the diurnal cycle is out of phase with the bias might result in an increase in error at that time. This would be then reflected in the unbiased component of the total error. The question whether the actual random errors are impacted by systematic error correction requires first including the diurnal cycle as well. This discussed in the future works section.

4.1.5 How does the performance of such online systematic error correction compare with the offline methods?

The AOC scheme outperformed the offline correction scheme at lead times longer than 1-day. The tested offline scheme applied different corrections at different forecast lead times calculated as the difference between the analysis and forecast for

that corresponding lead time. Though the offline corrected forecasts at short-time were better than online corrected in terms of bias, the offline correction leads to a higher increase in the unbiased component of total error.

4.2 Limitation of using Analysis Increments

Although the scheme successfully improved the global forecast for lead times shorter than 1 day and in the tropics for even longer lead times outperforming the offline correction scheme, it is crucial to acknowledge the assumptions made and limitations of this scheme.

Among several choices for verification data, the GDAS analysis best fits our purpose of computing corrections and using them to correct model tendency equation. GDAS analysis is generated using the same dynamical model and hence dynamically consistent with the forecasts. Further, it is available at the required grid location and at the required time avoiding the complexities related to spatiotemporal interpolation. However, AIs might not always indicate model deficiencies. Averaged analysis increment not only contain information about the model bias but also have contributions from averaged analysis errors and model error due to inaccuracies in the initial conditions.The collective contribution of the latter two terms has been referred to the β term through out the text. For a non-linear system, with Gaussian background and observation errors, highly reliable and sufficient observation with unbiased errors and negligible non-linear terms, AIs averaged over a large sample would have negligible β term and thus provide a good approximate of model

systematic errors. And thus AOC would lead to reduction in model biases.

However, given all the assumptions made above do not always hold β term cannot be neglected. Though it's a complex and involved problem to quantify the cumulative impact of each factor contributing to the β term, following are some cases where estimating model biases using AIs would not be accurate:

1. **When large observation related biases are present.** A major limitation of this scheme is its inability to account for biases in the analysis which result from the observations related errors during data assimilation. These can be from measurement errors, errors in the observation operator or representative errors. For example, AIs indicated that the amplitude of the seasonal biases declined in 2015, especially over the ocean. The large seasonal bias before 2015 was not completely due to model deficiencies but also arose from bias in prescribed SSTs and a problem with observational assimilation.

2. **When analysis contains significant portion of model biases.** In cases apart from where $\mathbf{R} << \mathbf{B}$, analysis would retain some portion of the model bias from the data assimilation process. This would lead to underestimation of model biases.

3. **Observation and background error statistics are non-Gaussian**. In this case, the analysis computed is not optimal and hence not the best estimate of truth. This would increase analysis error that would further increase the β term reducing the accuracy of model bias estimates.

4. **Significant higher order non-linear term.** Higher non-linearity in the ob-

servation operator or model error growth would introduce errors in the analysis and background respectively. Though the cumulative impact is hard to deduce, it would impact the bias estimates considerably. For example, the effectiveness of AOC scheme reduces after 1 day as the nonlinear error growth begins in the model.

5. **Using insufficient training period.** An insufficient training period would introduce sampling errors in the bias estimate. In this study, the smallest training period used is 7 days. This has been shown to be sufficient in observation dense areas by several past studies. However, in absence of sufficient reliable observations our estimate of model bias might contain some sampling biases.

6. **Observation and model bias corrections are not independent.** As mentioned earlier, several bias correction schemes correct observation biases using analysis as verifying truth. In case the analysis is biased to model deficiencies only, the observations would end up being nudged towards the incorrect model. Then if analyses using these observations are used to estimate model bias, it would result in underestimation of model bias. The problem is further complicated when biases are present in both observations and model and cycling comes into the picture.

Though several studies have attempted to address this problem (refer to section 2.4.1 for more details), these methods have been tested in an OSSE framework where the structure of model and observation error is known. For example, Privé and

Errico (2013) ound that in presence of both the model and observation biases, the forecast error arose mainly due to model errors and dominated the error growth in first couple days. This, if true for the GFS model as well, would mean that AIs calculated during first 6-hr would be dominated by model biases rather than biases initial condition. They also stated that analysis increments may underestimate the bias in the model as analysis may contain errors similar to those in short-term forecasts. Hence, these estimates might provide a lower limit of the actual model error. However, in the real world, the truth is unknown and only estimated through analysis and reanalysis from different meteorological centers. One possible way to test the presence of observation bias is by investigating the impact on the online correction on the bias, since erroneously correcting the model for an observation bias should result in an increase of the AIs.

Though we dont have a truth in the real world to compare the GDAS analysis with, analyses from other operational centers or reanalyses provide some estimate of how much the state-of-art analysis and reanalysis agree with each other. Comparison of the GDAS analysis with the ECMWF operational analysis, ERA5, which is the ECMWF Reanallysis 5 using 4DVar, and NCEP's Reanalysis 2 (fig. 4.1 shows that apart from near southern Polar region, near surface and near tropical troposphere the temperature analysis agree within +/-0.2K. The GDAS temperature analysis seems to be about 1-1.5K colder than other analyses over the high latitudes in the southern hemisphere and warmer than the ECMWF and ERA5 analyses near surface, over the equator and the tropics by about 1-1.5K. The GDAS temperatures below the tropopause also seem to be warmer by about 0.4K than other analyses/reanalyses.

It was shown in Fig. 2.4 that the GFS forecast was about 0.5 K warmer than the analysis in the latitude band 30-60S. Assuming the ERA5 and the ECMWF analyses to be more accurate, 6-hr AIs underestimate the model warm bias by about 1-1.5K.

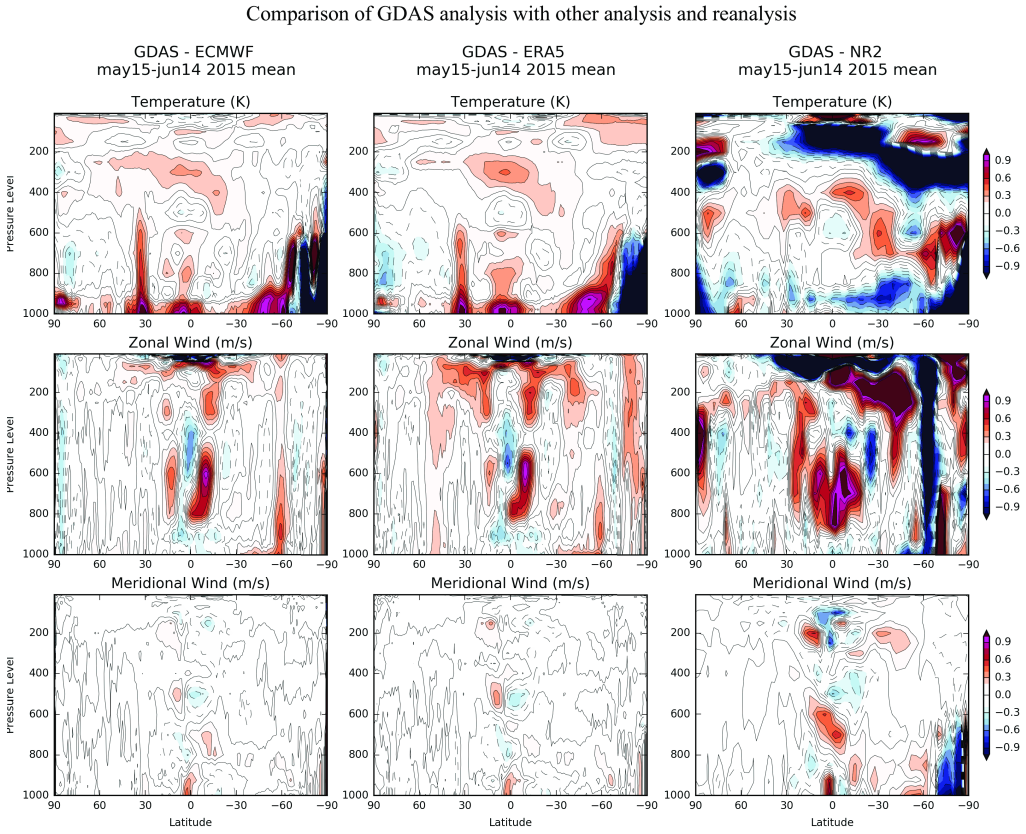

Figure 4.1: Comparison of zonal mean deviations of GDAS analysis from ECMWF operational analysis (left), ERA5 (center), NCEP Reanalysis 2 (right) averaged from May 15-June 14, 2015 for Temperature (K) (top), U-wind (middle) and the V-wind (bottom).

4.3 Future Directions

The very first experiments continuing this work should be aimed at correcting diurnal cycle errors. We found that application of online correction resulted in slightly increased unbiased component of error. As diurnal and semi-diurnal errors contribute significantly to the total bias, correcting only the mean bias should not be enough. The diurnal and semi-diurnal biases dominate the higher frequencies (sub-monthly) in GFS. As these are reproduced by four eigenmodes out of 120 modes, experiments using the DKM07 low dimensional approach, detailed in section 3.2 can be implemented.

So far we have analyzed the impact of online correction on the variables corrected. We suggest analyzing the impact of correcting the tested variables on other model variables like rainfall to gauge if the scheme improves the model overall or just the corrected variables. Once the adaptive online correction scheme that corrects both, the time mean as well as the periodic components of the state-independent errors, is tested with the GFS, the same can be used to improve the forecasts in a coupled model.

Apart from the model forecasts, model deficiencies also lead to divergence of the data assimilation system. It has been shown using simple but realistic SPEEDY model that online correction along with correction for random errors improved the LETKF analysis (Li et al., 2009). Any impact of online model correction on the analysis scheme would reflect in the bias estimates of cycled experiments. As the new analysis would be used at later times for verification and impact the bias cor-

rection estimates as well. Hence, before the AOC scheme can be tested for online implementation, it is important to understand the impact model correction has on analysis

The scheme has not yet been tested with cycled experiments which assimilate observations with the corrected forecasts. Similar to AIs, analysis minus observation is used to correct the observation related biases. This leads to observations being relaxed to the model forecasts. In order to implement this methodology for operational purposes, the interaction of observation and model bias correction schemes, both of which incorrectly assume bias free analysis, needs to be examined. We suggest testing this scheme with simple but realistic models in presence of observation and model biases using an OSSE framework first and then with operational scale models.

References

Allen, M., D. Frame, J. Kettleborough, and D. Stainforth (2006). Model error in weather and climate forecasting. *Predictability of Weather and Climate*, 391–427.

Alpert, J. C. and S. Saha (1989). Operational Systematic Error Correction for the NMC Operational Medium Range Forecast Model. Technical Report September.

Bauer, P., A. Thorpe, and G. Brunet (2015, 9). The quiet revolution of numerical weather prediction. *Nature 525*(7567), 47–55.

Bhargava, K., E. Kalnay, J. A. Carton, and F. Yang (2018, 2). Estimation of Systematic Errors in the GFS Using Analysis Increments. *Journal of Geophysical Research: Atmospheres 123*(3), 1626–1637.

Bjerknes, V. (1904). Das Problem der Wettervorhersage, betrachtet vom Der Standpunkte der Mechanik und Physik. *Meteorol. Z. 21*, 1–7.

Boukabara, S. A., T. Zhu, H. L. Tolman, S. Lord, S. Goodman, R. Atlas, M. Goldberg, T. Auligne, B. Pierce, L. Cucurull, M. Zupanski, M. Zhang, I. Moradi, J. Otkin, D. Santek, B. Hoover, Z. Pu, X. Zhan, C. Hain, E. Kalnay, D. Hotta, S. Nolin, E. Bayler, A. Mehra, S. P. F. Casey, D. Lindsey, L. Grasso, V. K. Kumar, A. Powell, J. Xu, T. Greenwald, J. Zajic, J. Li, J. Li, B. Li, J. Liu, L. Fang, P. Wang, T.-C. Chen, S. A. Boukabara, T. Zhu, H. L. Tolman, S. Lord, S. Goodman, R. Atlas, M. Goldberg, T. Auligne, B. Pierce, L. Cucurull, M. Zupanski, M. Zhang, I. Moradi, J. Otkin, D. Santek, B. Hoover, Z. Pu, X. Zhan, C. Hain, E. Kalnay, D. Hotta, S. Nolin, E. Bayler, A. Mehra, S. P. F. Casey, D. Lindsey, L. Grasso, V. K. Kumar, A. Powell, J. Xu, T. Greenwald, J. Zajic, J. Li, J. Li, B. Li, J. Liu, L. Fang, P. Wang, and T.-C. Chen (2016, 12). S4: An O2R/R2O Infrastructure for Optimizing Satellite Data Utilization in NOAA Numerical Modeling Systems: A Step Toward Bridging the Gap between Research and Operations. *Bulletin of the American Meteorological Society 97*(12), 2359–2378.

Buehner, M., J. Morneau, and C. Charette (2013). Four-dimensional ensemble-variational data assimilation for global deterministic weather prediction. *Nonlin. Processes Geophys 20*, 669–682.

Carter, G. M., J. P. Dallavalle, and H. R. Glahn (1989, 9). Statistical Forecasts Based on the National Meteorological Center's Numerical Weather Prediction System. *Weather and Forecasting 4*(3), 401–412.

Chapman, W. L., J. E. Walsh, W. L. Chapman, and J. E. Walsh (2007, 2). Simulations of Arctic Temperature and Pressure by Global Coupled Models. *Journal of Climate 20*(4), 609–632.

Cheng, W. Y. Y., W. J. Steenburgh, W. Y. Y. Cheng, and W. J. Steenburgh (2007, 12). Strengths and Weaknesses of MOS, Running-Mean Bias Removal, and Kalman Filter Techniques for Improving Model Forecasts over the Western United States. *Weather and Forecasting 22*(6), 1304–1318.

Cui, B., Z. Toth, Y. Zhu, and D. Hou (2012, 4). Bias Correction for Global Ensemble Forecast. *Weather and Forecasting 27*(2), 396–410.

Dalcher, A. and E. Kalnay (1987, 10). Error growth and predictability in operational ECMWF forecasts. *Tellus A 39A*(5), 474–491.

Danforth, C. M. and E. Kalnay (2008a). Impact of online empirical model correction on nonlinear error growth. *Geophysical Research Letters 35*(24), 1–6.

Danforth, C. M. and E. Kalnay (2008b). Using Singular Value Decomposition to Parameterize State-Dependent Model Errors.

Danforth, C. M., E. Kalnay, and T. Miyoshi (2007). Estimating and correcting global weather model error. *Bulletin of the American Meteorological Society 88*(3), 303–304.

Dee, D. P. (2005a). Bias and data assimilation. *131*, 3323–3343.

Dee, D. P. (2005b, 10). Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society 131*(613), 3323–3343.

Dee, D. P. and S. Uppala (2009). Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. *QUARTERLY JOURNAL OF THE ROYAL METEOROLOGICAL SOCIETY Q. J. R. Meteorol. Soc 135*, 1830–1841.

Delle Monache, L., T. Nipen, Y. Liu, G. Roux, R. Stull, L. D. Monache, T. Nipen, Y. Liu, G. Roux, and R. Stull (2011, 11). Kalman Filter and Analog Schemes to Postprocess Numerical Weather Predictions. *Monthly Weather Review 139*(11), 3554–3570.

DelSole, T. and A. Y. Hou (1999, 11). Empirical Correction of a Dynamical Model. Part I: Fundamental Issues. *Monthly Weather Review 127*(11), 2533–2545.

DelSole, T., M. Zhao, and P. Dirmeyer (2009). A New Method for Exploring Coupled LandAtmosphere Dynamics. *Journal of Hydrometeorology 10*(4), 1040–1050.

DelSole, T., M. Zhao, P. a. Dirmeyer, and B. P. Kirtman (2008). Empirical Correction of a Coupled LandAtmosphere Model. *Monthly Weather Review 136*(1978), 4063–4076.

Faller, A. J. and C. E. Schemm (1977, 1). Statistical Corrections to Numerical Prediction Equations. II. *Monthly Weather Review 105*(1), 37–56.

Fan, Y. and H. van den Dool (2011, 6). Bias Correction and Forecast Skill of NCEP GFS Ensemble Week-1 and Week-2 Precipitation, 2-m Surface Air Temperature, and Soil Moisture Forecasts. *Weather and Forecasting 26*(3), 355–370.

Frei, T. (2009). Economic and social benefits of meteorology and climatology in Switzerland. *METEOROLOGICAL APPLICATIONS Meteorol. Appl.*

Gel, Y. R. (2007, 12). Comparative Analysis of the Local Observation-Based (LOB) Method and the Nonparametric Regression-Based Method for Gridded Bias Correction in Mesoscale Weather Forecasting. *Weather and Forecasting 22*(6), 1243–1256.

Glahn, H. R. and D. A. Lowry (1972). The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. *Journal of Applied Meteorology*, 1203–1211.

Greybush, S. J., R. J. Wilson, R. N. Hoffman, M. J. Hoffman, T. Miyoshi, K. Ide, T. McConnochie, and E. Kalnay (2012, 11). Ensemble Kalman filter data assimilation of Thermal Emission Spectrometer temperature retrievals into a Mars GCM. *Journal of Geophysical Research: Planets 117*(E11), n/a–n/a.

Hacker, J. P. and D. L. Rife (2007, 12). A Practical Approach to Sequential Estimation of Systematic Error on Near-Surface Mesoscale Grids. *Weather and Forecasting 22*(6), 1257–1273.

Hartmann, D. L., T. N. Palmer, and R. Buizza (1996, 8). Finite-Time Instabilities of Lower-Stratospheric Flow. *Journal of the Atmospheric Sciences 53*(15), 2129–2143.

Houtekamer, P. L., F. Zhang, P. L. Houtekamer, and F. Zhang (2016, 12). Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation. *Monthly Weather Review 144*(12), 4489–4532.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh (2007, 6). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena 230*(1-2), 112–126.

Johansson, . and S. Saha (1989). Simulation of Systematic Error Effects and Their Reduction in a Simple Model of the Atmosphere. *Monthly Weather Review 117*(8), 1658–1675.

Jung, T. (2005). Systematic errors of the atmospheric circulation in the ECMWF forecasting system. *Quarterly Journal of the Royal Meteorological Society 131*(607), 1045–1073.

Jung, T. and A. Tompkins (2003a). Forecasting System. (October).

Jung, T. and A. Tompkins (2003b). Systematic Errors in the ECMWF Forecasting System. Technical Report October, Technical Memorendum.

Kalnay, E. (2003). *Atmospheric modeling, data assimilation, and predictability.*

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, R. Jenne, D. Joseph, E. Kalnay, M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne, and D. Joseph (1996, 3). The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society 77*(3), 437–471.

Kleist, D. T. and K. Ide (2015, 2). An OSSE-Based Evaluation of Hybrid VariationalEnsemble Data Assimilation for the NCEP GFS. Part II: 4DEnVar and Hybrid Variants. *Monthly Weather Review 143*(2), 452–470.

Kleist, D. T., D. F. Parrish, J. C. Derber, R. Treadon, W.-S. Wu, and S. Lord (2009, 12). Introduction of the GSI into the NCEP Global Data Assimilation System. *Weather and Forecasting 24*(6), 1691–1705.

Klinker, E. and P. D. Sardeshmukh (1992). The Diagnosis of Mechanical Dissipation in the Atmosphere from Large-Scale Balance Requirements.

Klocke, D. and M. J. Rodwell (2014, 1). A comparison of two numerical weather prediction methods for diagnosing fast-physics errors in climate models. *Quarterly Journal of the Royal Meteorological Society 140*(679), 517–524.

Leith, C. E. (1978). Objective Methods for Weather Prediction. *Annual Review Fluid Mechanics*, 107–128.

Li, H., E. Kalnay, T. Miyoshi, and C. M. Danforth (2009). Accounting for Model Errors in Ensemble Data Assimilation. *Monthly Weather Review 137*(10), 3407–3419.

Lien, G.-Y., T. Miyoshi, E. Kalnay, G.-Y. Lien, T. Miyoshi, and E. Kalnay (2016, 2). Assimilation of TRMM Multisatellite Precipitation Analysis with a Low-Resolution NCEP Global Forecast System. *Monthly Weather Review 144*(2), 643–661.

Lorente-Plazas, R. and J. P. Hacker (2017, 3). Observation and model bias estimation in the presence of either or both sources of error. *Monthly Weather Review*, 16–0273.

Mao, Q., R. T. McNider, S. F. Mueller, H.-M. H. Juang, Q. Mao, R. T. McNider, S. F. Mueller, and H.-M. H. Juang (1999, 4). An Optimal Model Output Calibration Algorithm Suitable for Objective Temperature Forecasting. *Weather and Forecasting 14*(2), 190–202.

Marzban, C. (2003, 6). Neural Networks for Postprocessing Model Output: ARPS. *Monthly Weather Review 131*(6), 1103–1111.

Marzban, C., S. Sandgathe, E. Kalnay, C. Marzban, S. Sandgathe, and E. Kalnay (2006, 2). MOS, Perfect Prog, and Reanalysis. *Monthly Weather Review 134*(2), 657–663.

Molteni, F. (2003). Atmospheric simulations using a GCM with simplified physical parametrizations. I: model climatology and variability in multi-decadal experiments. *Climate Dynamics 20*(2), 175–191.

Murphy, A. H. (1988, 12). Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient. *Monthly Weather Review 116*(12), 2417–2424.

Orrell, D., L. Smith, J. Barkmeijer, and T. N. Palmer (2001). Model error in weather forecasting. *Nonlinear Processes in Geophysics 8*(June), 357–371.

Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke (2002, 3). A Local Ensemble Kalman Filter for Atmospheric Data Assimilation.

Pauwels, V. R. N., G. J. M. De Lannoy, H.-J. Hendricks Franssen, H. Vereecken, and Pauwels V De Lannoy G Hendricks Franssen H Vereecken H (2013, 9). Simultaneous estimation of model state variables and observation and forecast biases using a two-stage hybrid Kalman filter. *Hydrology and Earth System Sciences 17*(9), 3499–3521.

Perrels, A., T. Frei, F. Espejo, L. Jamin, and A. Thomalla (2013). Socio-economic benefits of weather and climate services in Europe. *Adv. Sci. Res 10*, 2012.

Privé, N. C. and R. M. Errico (2013). The role of model and initial condition error in numerical weather forecasting investigated with an observing system simulation experiment. *Tellus A: Dynamic Meteorology and Oceanography 65*.

Privé, N. C. and R. M. Errico (2015). Spectral analysis of forecast error investigated with an observing system simulation experiment. *Tellus A: Dynamic Meteorology and Oceanography 67*.

Reynolds, R. W. and T. M. Smith (1994, 6). Improved Global Sea Surface Temperature Analyses Using Optimum Interpolation. *Journal of Climate 7*(6), 929–948.

Saha, S. (1992). Response of NMC MRF Model to Systematic-Error Correction within Integration. *Monthly Weather Review 120*, 345–360.

Schemm, C. E., D. A. Unger, and A. J. Faller (1981, 1). Statistical Corrections to Numerical Predictions III. *Monthly Weather Review 109*(1), 96–109.

Schemm, J. K. E. and A. J. Faller (1986, 12). Statistical Corrections to Numerical Predictions. Part IV. *Monthly Weather Review 114*(12), 2402–2417.

Sela, J. (2009). The Implementation of the Sigma Pressure Hybrid Cordinates into the GFS. Technical report, National Oceanic and Atmospheric Administation National Weather Service National Centers for Environmental Prediction.

Stensrud, D. J. and N. Yussouf (2003, 10). Short-Range Ensemble Predictions of 2-m Temperature and Dewpoint Temperature over New England. *Monthly Weather Review 131*(10), 2510–2524.

Stensrud, D. J. and N. Yussouf (2005, 10). Bias-corrected short-range ensemble forecasts of near surface variables. *Meteorological Applications 12*(03), 217.

Thiébaux, J., E. Rogers, W. Wang, B. Katz, J. Thiébaux, E. Rogers, W. Wang, and B. Katz (2003, 5). A New High-Resolution Blended Real-Time Global Sea Surface Temperature Analysis. *Bulletin of the American Meteorological Society 84*(5), 645–656.

Vannitsem, S. and Z. Toth (2002). Short-Term Dynamics of Model Errors. *Journal of the Atmospheric Sciences 59*(17), 2594–2604.

Wang, X., D. Parrish, D. Kleist, J. Whitaker, X. Wang, D. Parrish, D. Kleist, and J. Whitaker (2013, 11). GSI 3DVar-Based EnsembleVariational Hybrid Data Assimilation for NCEP Global Forecast System: Single-Resolution Experiments. *Monthly Weather Review 141*(11), 4098–4117.

Wilson, L. J. and M. Vallée (2002, 4). The Canadian Updateable Model Output Statistics (UMOS) System: Design and Development Tests. *Weather and Forecasting 17*(2), 206–222.

Woodcock, F. and C. Engel (2005, 2). Operational Consensus Forecasts. *Weather and Forecasting 20*(1), 101–111.

Wu, W.-S., R. J. Purser, D. F. Parrish, W.-S. Wu, R. J. Purser, and D. F. Parrish (2002, 12). Three-Dimensional Variational Analysis with Spatially Inhomogeneous Covariances. *Monthly Weather Review 130*(12), 2905–2916.

Xue, H., X. Shen, and J. Chou (2015a, 11). An online model correction method based on an inverse problem: Part IIsystematic model error correction. *Advances in Atmospheric Sciences 32*(11), 1493–1503.

Xue, H., X. Shen, and J. Chou (2015b, 10). An online model correction method based on an inverse problem: Part IModel error estimation by iteration. *Advances in Atmospheric Sciences 32*(10), 1329–1340.

Yang, F. (2015). Decomposition of Model Forecast Errors: Methodology and Application. *American Geophysical Union, Fall Meeting 2015, abstract id. A21E-0190*.

Yang, F., C. Andrew, T. Russ, and D. John (2013). On Improving GFS Forecast Skills in the Southern hemisphere: Ideas and Preliminary Results. In *NCEP-EMC Global Modelling Branch Bi-Weekly Briefing*.

Yang, X., T. DelSole, and H.-L. Pan (2008, 12). Empirical Correction of the NCEP Global Forecast System. *Monthly Weather Review 136*(12), 5224–5233.

Yussouf, N. and D. J. Stensrud (2006, 11). Prediction of Near-Surface Variables at Independent Locations from a Bias-Corrected Ensemble Forecasting System. *Monthly Weather Review 134*(11), 3415–3424.

Zheng, Y., T. Shinoda, L.-J. Lin, and G. N. Kiladis (2011). Sea Surface Temperature Biases under the Stratus Cloud Deck in the Southeast Pacific Ocean in 19 IPCC AR4 Coupled General Circulation Models. *Journal of Climate*.

Zhu, Y., J. C. Derber, R. J. Purser, B. A. Ballish, J. Whiting, Y. Zhu, J. C. Derber, R. J. Purser, B. A. Ballish, and J. Whiting (2015, 9). Variational Correction of Aircraft Temperature Bias in the NCEPs GSI Analysis System. *Monthly Weather Review 143*(9), 3774–3803.

Zhu, Y., E. Liu, R. Mahajan, C. Thomas, D. Groff, P. Van Delst, A. Collard, D. Kleist, R. Treadon, J. C. Derber, Y. Zhu, E. Liu, R. Mahajan, C. Thomas, D. Groff, P. V. Delst, A. Collard, D. Kleist, R. Treadon, and J. C. Derber (2016, 12). All-Sky Microwave Radiance Assimilation in NCEPs GSI Analysis System. *Monthly Weather Review 144*(12), 4709–4735.